



## OPEN ACCESS

## EDITED BY

Vickramjeet Singh,  
Dr. B. R. Ambedkar National Institute of  
Technology Jalandhar, India

## REVIEWED BY

Yilun Shang,  
Northumbria University, United Kingdom  
Imran Khan,  
Sultan Qaboos University, Oman

## \*CORRESPONDENCE

Muhammad Waheed Rasheed,  
✉ waheedrasheed461@gmail.com

RECEIVED 07 February 2024

ACCEPTED 22 March 2024

PUBLISHED 18 April 2024

## CITATION

Alam SM, Nigar N, Rasheed MW and Amin L  
(2024), Analyzing the role of reducible  
molecular descriptors and thermodynamic  
aspects of anti-tuberculosis drugs via  
QSPR study.  
*Front. Phys.* 12:1383216.  
doi: 10.3389/fphy.2024.1383216

## COPYRIGHT

© 2024 Alam, Nigar, Rasheed and Amin. This is  
an open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](#). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic practice.  
No use, distribution or reproduction is  
permitted which does not comply with these  
terms.

# Analyzing the role of reducible molecular descriptors and thermodynamic aspects of anti-tuberculosis drugs via QSPR study

Sajid Mahboob Alam<sup>1</sup>, Niat Nigar<sup>1</sup>,  
Muhammad Waheed Rasheed<sup>2\*</sup> and Laiba Amin<sup>2</sup>

<sup>1</sup>School of Mathematics, Minhaj University Lahore, Lahore, Pakistan, <sup>2</sup>Department of Mathematics, Division of Science and Technology, University of Education, Lahore, Pakistan

*Mycobacterium tuberculosis* is an organism that causes tuberculosis (TB), a common infectious disease that has a high death and morbidity rate. Topological indices are mathematical tools used to describe the structural properties of molecules or networks. They provide a quantitative measure of the connectivity and complexity of a system, and play a crucial role in numerous areas such as biochemistry and bioinformatics. The purpose of topological indices is to simplify complex structures into numerical values that can be easily analyzed and compared. QSPR modeling is a technique in chemistry that relates the structure of a chemical compound to its physical or chemical properties. It is used to predict properties like boiling points, solubilities, toxicities, and even biological activities of compounds. This saves time, resources, and enables researchers to make informed decisions in drug discovery, material science, and many other areas. In this study, we conducted an analysis of several drugs used for the treatment of tuberculosis. We focused on computing the reducible topological indices based on their degrees. Several techniques and approaches are employed. To perform calculations, we used edge partition methodology, analytical techniques, theoretical graph utilities, and degree counting method. Additionally, we examined six physicochemical properties of these drugs. To establish quantitative structure-property relationship models and evaluate their effectiveness, we employed linear, quadratic, and logarithmic regression analysis. By analyzing the reducible topological indices and physicochemical properties, we aimed to gain a deeper understanding of the drugs' characteristics and their potential impact on tuberculosis treatment. This study established a significant relationship between the defined indices with two key properties: molar mass and collision cross section. The correlation coefficients for molar mass range from 0.7 to 0.9, while the collision cross section range from 0.8 to 0.9. These results demonstrate a strong association between the indices and the properties under investigation. Furthermore, it is worth noting that both molar mass and collision cross section satisfy the requirements for *p*-value and *F*-test value across all indices. This indicates the statistical significance of the observed correlations and the reliability of our findings.

## KEYWORDS

tuberculosis drugs, degree-based reducible topological indices, QSPR analysis, regression models, correlation coefficients (*r*)

## 1 Introduction

A dreadful disease in which a conflict between the soul and the body is so gradual, calm, and solemn, with such a sure outcome, that the mortal half rots away day by day and grain by grain. A disease that “sometimes moves in enormous steps and sometimes at a tardy, sluggish pace, but, slow or quick, is always sure and certain.” (Nicholas Nickleby; Charles Dickens). The things Charles Dickens said are still true today [1].

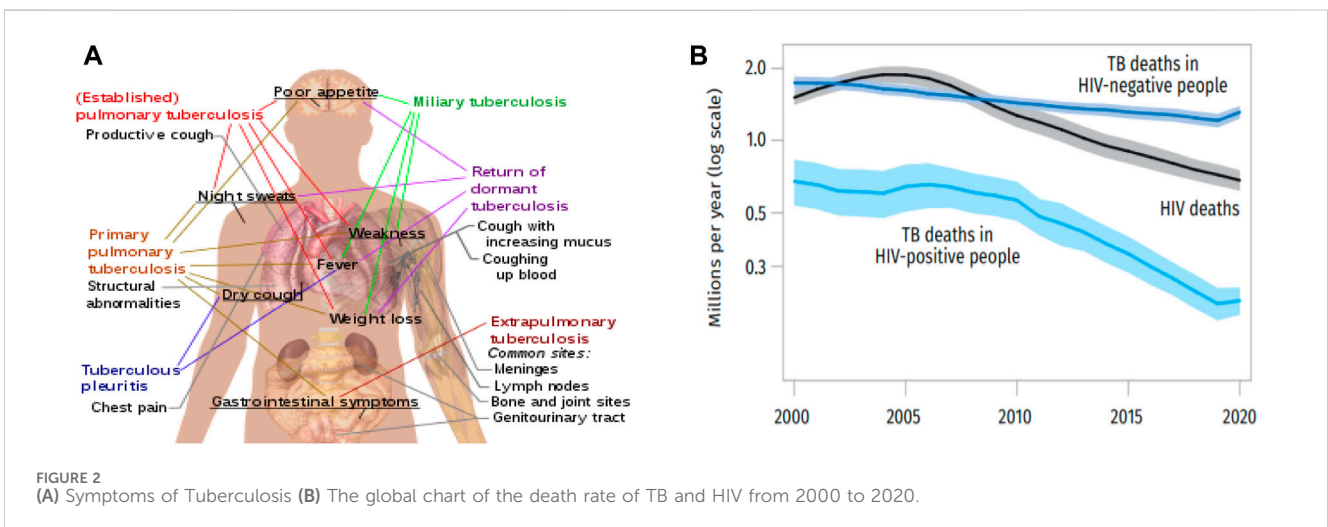
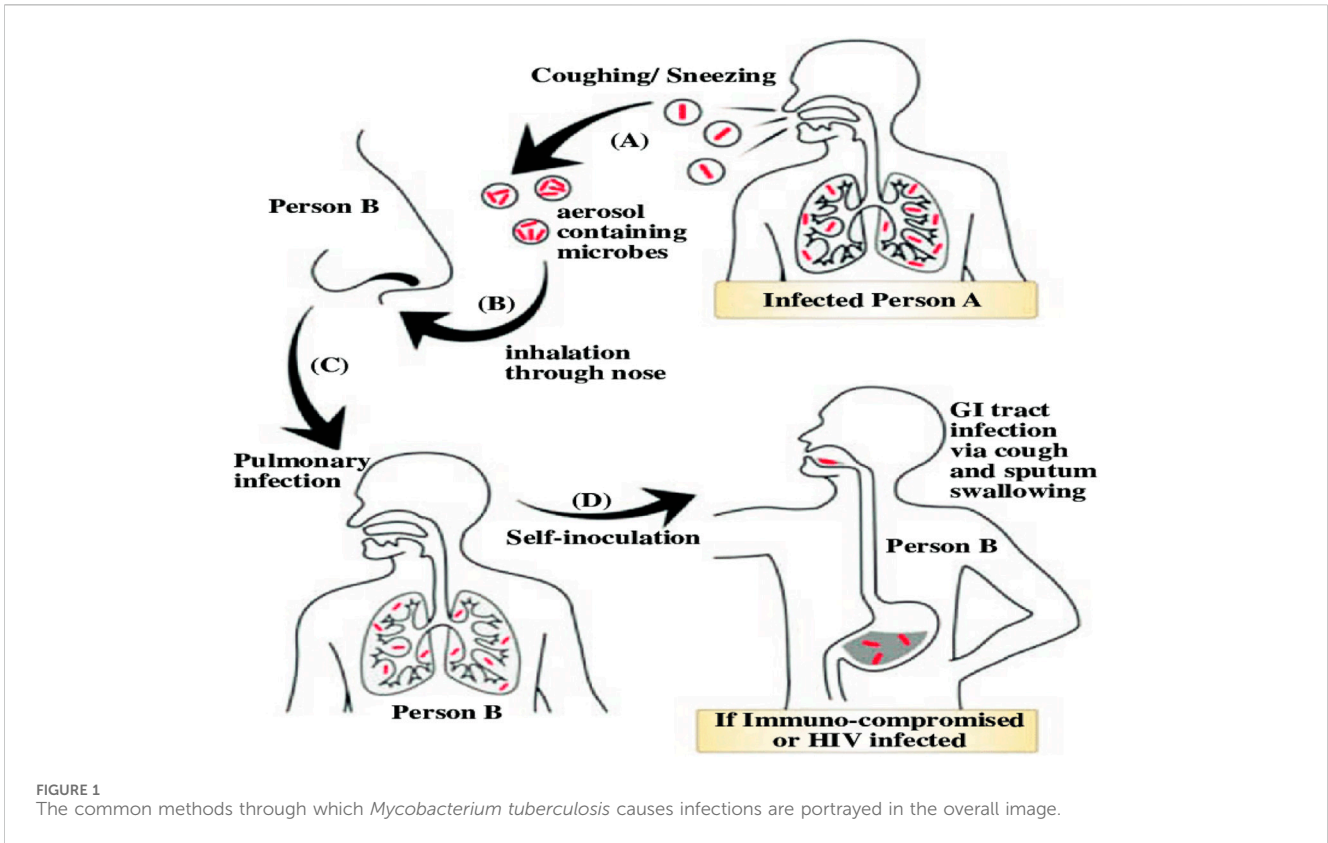
Bacteria called *Mycobacterium tuberculosis* (MTB) are typically responsible for the infectious disease known as tuberculosis (TB). Typically, tuberculosis primarily targets the lungs; nevertheless, it has the ability to impact various other regions of the body. In cases where there are no symptoms, it is referred to as latent tuberculosis [2]. Active disease occurs in about 10 percent of latent infections and, when untreated, results in a mortality rate of approximately 50 percent among those affected. TB is a global disease, although the majority of cases occur in low and middle-income nations. There are eight countries, namely, Bangladesh, China, India, Indonesia, Nigeria, Pakistan, Philippines, and South Africa, where approximately half of all individuals with TB can be located. It is believed that approximately 25 percent of the world's population carries a TB infection; however, the majority of individuals do not progress to acquire TB disease, and a small percentage may be able to overcome the infection. It cannot be spread by those who are infected but not (yet) unwell with the illness. The lifetime chance of developing tuberculosis is 5–10 percent for those with TB bacterium infection. Individuals with weakened immune systems, including those with HIV, diabetes, malnourishment, or tobacco use, are more susceptible to illness. People who are most likely to contract tuberculosis (TB) often fall into two distinct categories:

1. Those who have recently become infected with TB bacteria.
2. Those with weakened immune systems due to underlying medical conditions.

People who have active tuberculosis in their lungs cough, spit, speak, or sneeze can transmit the disease to others through the air. Latent tuberculosis patients do not transmit the illness. Figure 1 represented a general overview of the prevalent ways that *Mycobacterium tuberculosis* infections occur. In this illustration, the bacterium is released by an infected person (A) through sneezing or coughing, at which point the pathogen exits from person (A)'s body and comes as an aerosol. At B point, a healthy individual (B) inhales particles in the air containing MTB. The next stage is typically a pulmonary infection when the virus reaches the effective titre in person (B). When someone self-inoculates, the infection travels from the lung to the gastrointestinal tract. Because it occurs when person (B), who has pulmonary tuberculosis, swallows his own contaminated sputum, the term “self-inoculation” is employed.

TB illness symptoms vary depending on the location of the TB bacteria's growth within the body. The lungs are where TB germs typically grow called pulmonary TB. Chest pain, coughing up blood or sputum (phlegm from deep inside the lungs), and a severe cough lasting 3 weeks or longer are some of the signs of tuberculosis (TB) sickness in the lungs. Additional indications of tuberculosis illness include weakness or exhaustion, appetite loss, weight loss, chills, fever, and night sweats. TB illness symptoms vary depending on the affected area of the body. Individuals with latent tuberculosis infection do not exhibit any symptoms, do not feel ill, and are unable to infect others. Other types of tuberculosis (TB) are caused when the infection travels outside the lungs in 15–20 percent of active patients. All of these conditions are referred to as extrapulmonary tuberculosis [3, 4]. Extrapulmonary tuberculosis (TB) is more common in young children and those with weaker immune systems. This affects people who possess HIV in over 50 percent of situations. Tuberculosis can affect more than just the lungs. It can spread to different parts of the body, leading to extrapulmonary infections. Notable sites of infection include the pleura, which is the lining around the lungs, resulting in tuberculous pleurisy. The central nervous system can also be affected, causing tuberculous meningitis. The lymphatic system, responsible for immune function, can be infected, leading to scrofula of the neck. The genitourinary system, involving the reproductive and urinary organs, can also be targeted, resulting in urogenital tuberculosis. Additionally, tuberculosis can affect the bones and joints, causing Pott disease of the spine. These are just a few examples of the various extrapulmonary infection sites associated with tuberculosis. Figure 2A presented the main symptoms of variants and stages of tuberculosis.

It was estimated in 2018 that 25 percent of people on the planet may be subconsciously infected with tuberculosis. Every year, approximately 1 percent of the population contracts a new infection. After COVID-19, TB is the second most common infectious disease-related cause of mortality in 2020, with an anticipated 10 million cases of active TB and 1.5 million people dying. As of 2018, the majority of tuberculosis (TB) cases were reported in three regions: South-East Asia (44%), Africa (24%), and the Western Pacific (18%). Interestingly, more than half of all cases were diagnosed in just seven countries, namely, India (27%), China (9%), Indonesia (8%), the Philippines (6%), Pakistan (6%), Nigeria (4%), and Bangladesh (4%) [5]. These statistics shed light on the global distribution of TB, highlighting the concentration of cases in specific regions and countries. By 2021, there were only about 2% more new cases per year on average. When it comes to tuberculin testing, almost 80% of individuals in several Asian and African nations test positive, compared to 5%–10% of Americans. Humans have had tuberculosis from ancient times [6]. More than 2 million people globally per year pass away from tuberculosis (TB), the most common infectious disease-related cause of mortality. Around the world, one in three people, or 2-3 billion people, are estimated to be infected with *Mycobacterium tuberculosis* (*M. tuberculosis*), with a lifetime risk of active TB disease of 5%–15%. An estimated 9.6 million individuals were diagnosed with TB in 2014, and 1.5 million died from it, including 1.1 million HIV-negative people and 400,000 people living with HIV. Although TB exists in every country, the majority of TB patients live in low- and middle-income nations, particularly in areas like Sub-Saharan Africa and South East Asia [7]. Since 2000 until 2020, Figure 2B illustrates the worldwide trends in estimated death counts attributed to TB and HIV.



Antibiotics are used in the treatment of tuberculosis. Both the illness and infection from tuberculosis should be treated. The most often prescribed antibiotics include Lumefrantrine, Mefloquine, Piperaquine, Isoniazid, and Primaquine, among others. These drugs must be used consistently for 4–6 months in order to be effective. It is risky to quit taking the drugs too soon or without consulting a doctor. This may cause TB that is still alive to develop medication resistance.

The field of mathematics known as “graph theory” is devoted to the study of graphs’ characteristics and uses. A graph is a type of mathematical structure made up of edges (links) connecting nodes (vertices). These nodes and edges can be represented a numerous real-world scenarios, including social networks, transport systems, and computer networks. Graph theory offers a structure for using graph representations to analyze and solve difficult issues. The concept of the degree of a node in graph theory identifies the quantity of edges that are connected to that node. In order to grasp the connectivity and structure of a graph, it is crucial to understand this concept. An important area of study in graph theory is graph coloring, where colors are designated to nodes within a graph to prevent neighboring nodes from

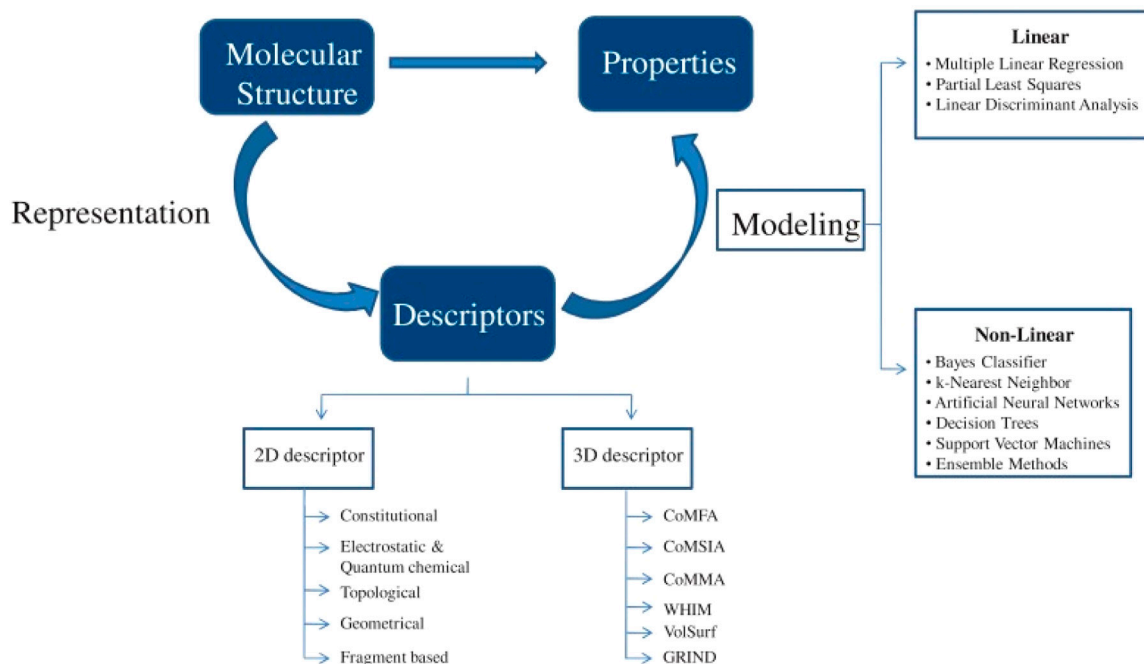


FIGURE 3 General methodology of a QSPR study.

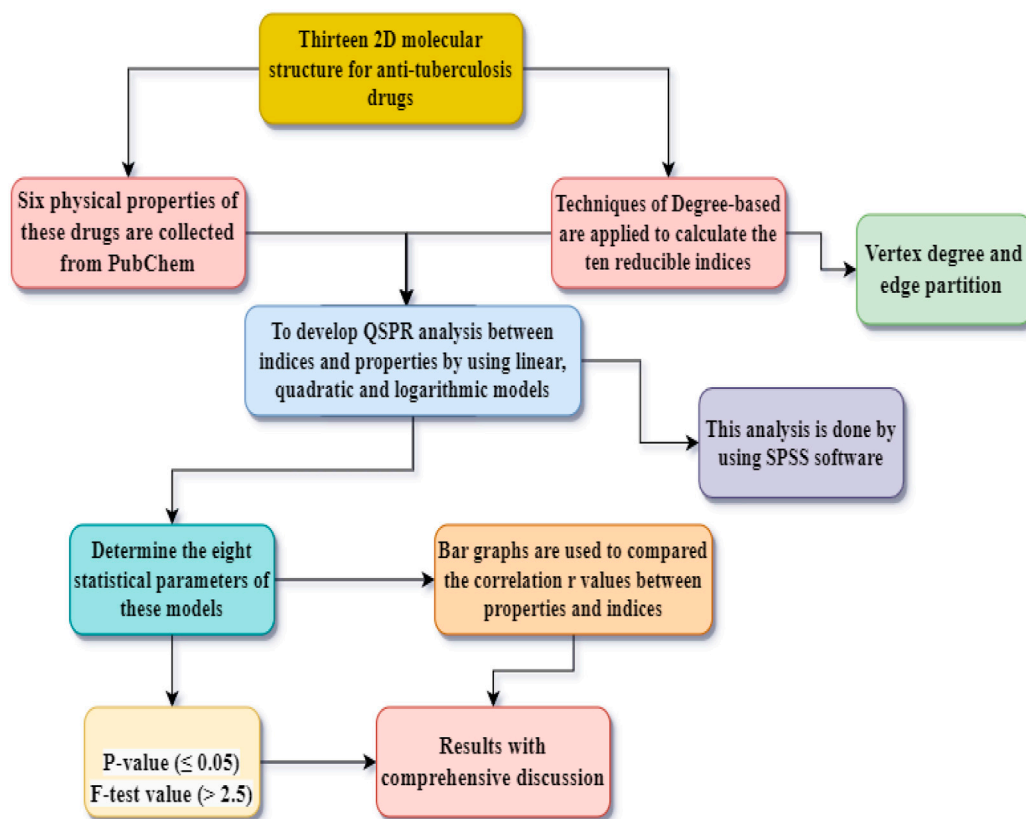


FIGURE 4 Flowchart of this manuscript.

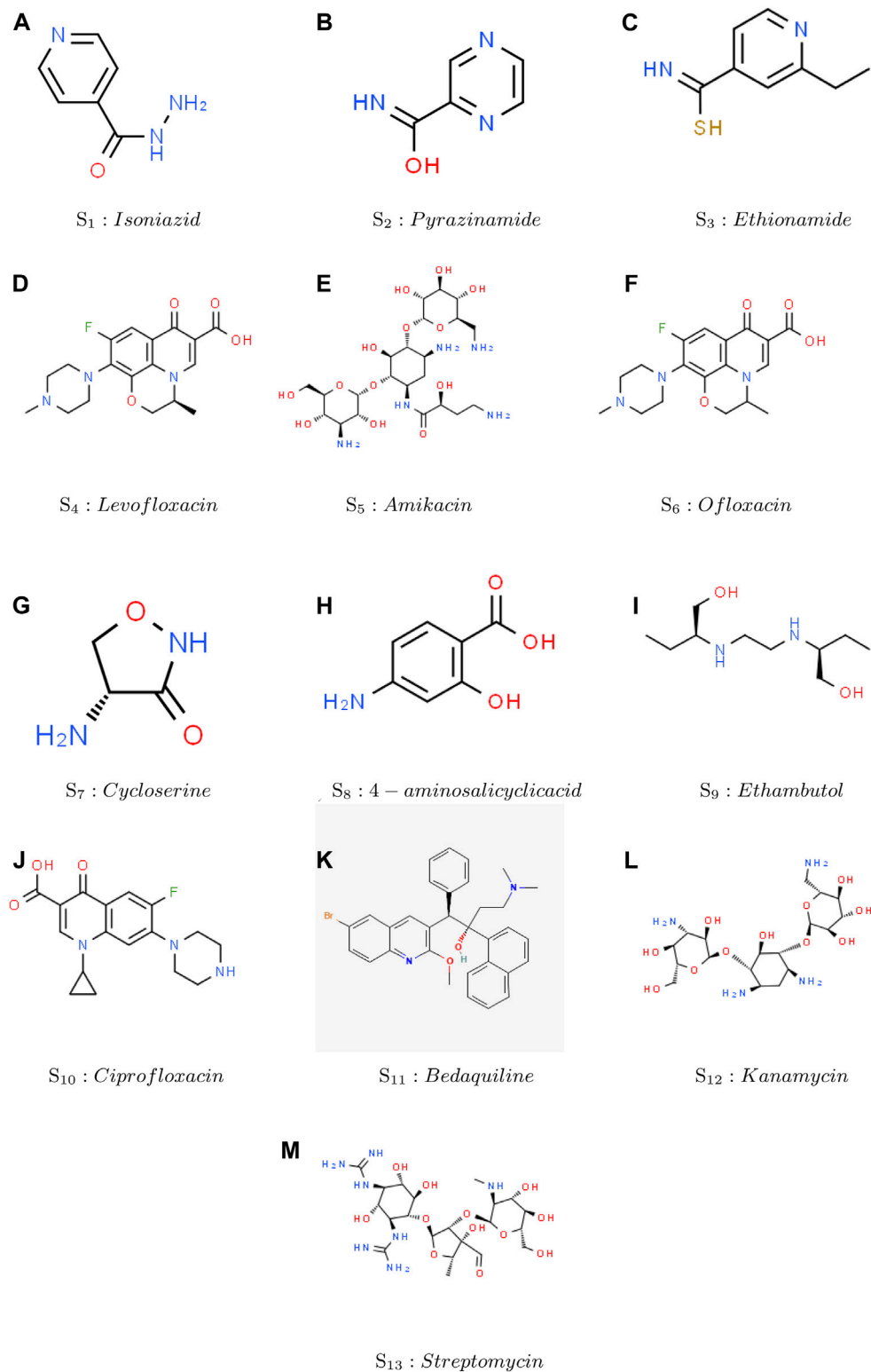


FIGURE 5  
Thirteen Chemical structure of drugs used to treat Tuberculosis disease.

sharing the same color. This concept is helpful in the realms of scheduling, resource allocation, and solving challenges related to map coloring. Graph theory plays a critical role in computer science, operations research, and biology due to its diverse range of practical applications. The idea of graph algorithms is crucial in solving problems related to network routing, social media analysis, and data visualization in computer science. Applying graph theory in operations research leads to the optimization of transportation networks, supply chains, and

TABLE 1 Thirteen drugs related to tuberculosis disease with their calculated reducible topological indices.

Drug name	$RM_1(G)$	$RM_2(G)$	RR(G)	RSC(G)	$RHM_1(G)$	$RHM_2(G)$
Isoniazid	100	244.4445	77.2999	60.1129	1055.5555	7067.9012
Pyrazinamide	66	144	31.39	21.38	648	3179.25
Ethionamide	121	315.94	57.60	36.20	1418.39	10958.16
Levofloxacin	676	3736.78	321.87	138.80	16937.56	573688.23
Amikacin	1,600	1444.44	754.08	256.33	66755.56	6401975.31
Ofloxacin	676	3736.78	321.87	138.80	16937.56	573688.23
Cycloserine	59.5	78.94	23.13	18.39	362.06	996.71
4-aminosalicylic acid	121	299.14	56.19	36.16	1425.11	9410.46
Ethambutol	196	735	94.89	50.05	3168.67	52021.67
Ciprofloxacin	576	2,960	277.41	123.84	13,024	376064
Bedaquiline	1,369	11123.13	662.38	235.43	48143.16	3525982.29
Kanamycin	1,089	7,986	511.81	192.98	37207.5	2349880.5
Streptomycin	1,444	11712.44	664.80	239.42	56275.89	4292548.49
RS(G)	RF(G)		$RG_1(G)$		$RG_2(G)$	
77.7778	566.6666		344.4445		2685.1851	
72	360		210		1,431	
154.61	665.5		469.13		3869.76	
1990.44	9,464		4412.78		97481.70	
8977.78	37866.66		16044.44		623847.41	
1990.44	9,464		4412.78		97481.70	
46.28	204.17		127.94		597.07	
228.56	826.83		420.14		3635.60	
228.67	1698.67		931		12576.67	
1,184	7,104		3,536		69,248	
3650.67	25896.92		12492.13		405399.88	
5263.5	21235.5		9,075		289492.5	
9426.11	32,851		13156.44		482670.37	

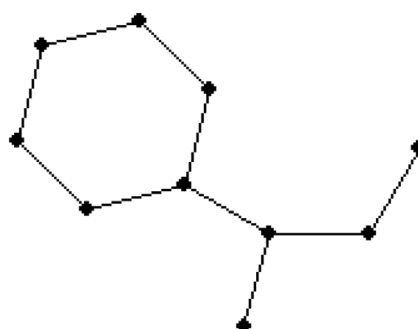
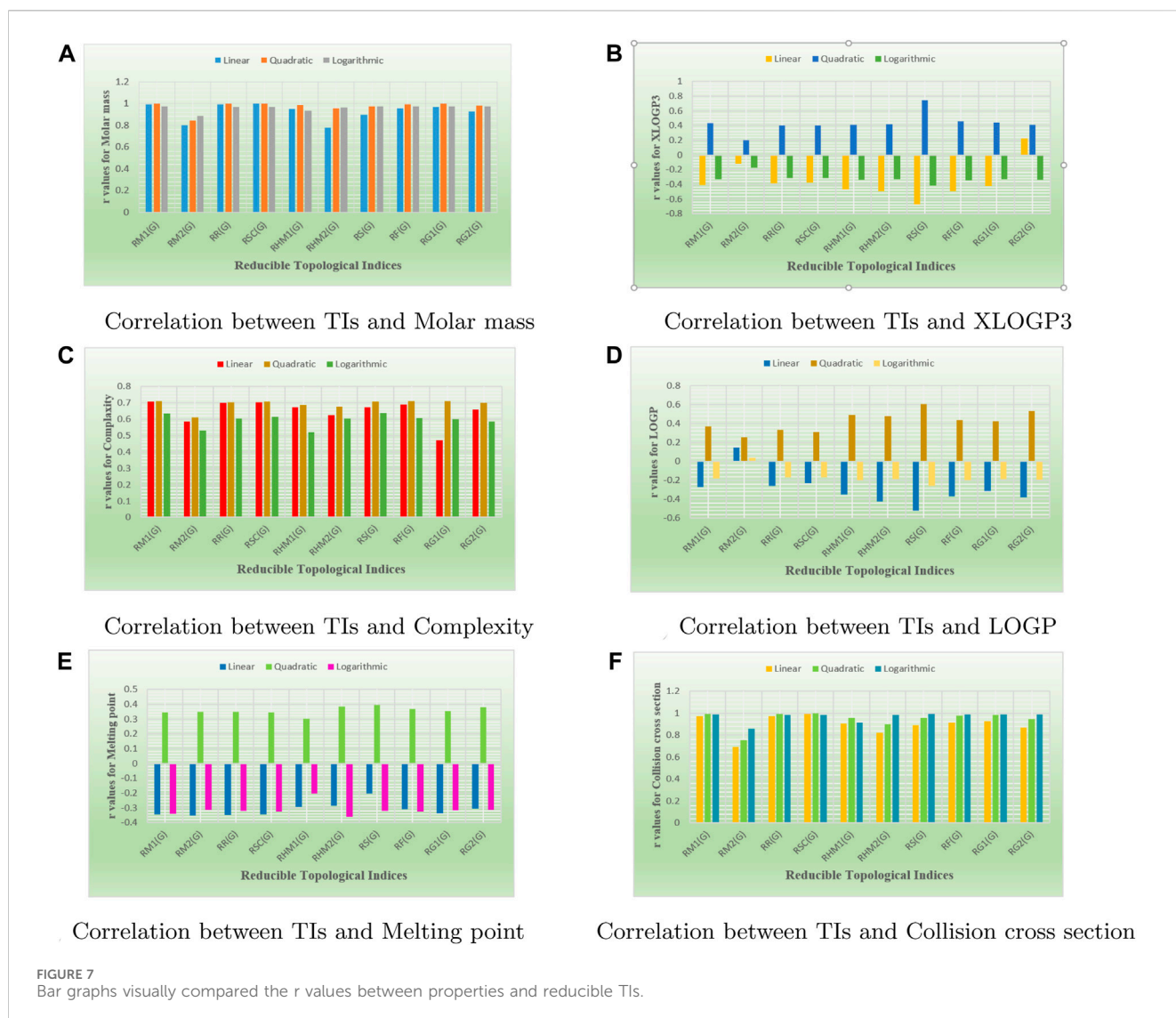


FIGURE 6 Molecular structure of the Isoniazid drug.



communication systems. The application of graph theory in the field of biology enables researchers to map out the connections among genes, proteins, and diseases, resulting in breakthroughs in network biology and personalized medicine. In general, graph theory is an effective tool that keeps discovering new uses in a variety of scientific and commercial fields.

The study of mathematical models of chemical phenomena using graph theory is known as chemical graph theory. Chemical graph theory has a subfield called topological indices that correlates specific physico-chemical characteristics of the underlying chemical molecule. Numerous publications on topological indices have been published to date, numbering in the hundreds [8]. A topological index is a function called “Top” from a set of finite simple graphs to the set of real numbers, where “ $\cong$ ” is the property that  $\text{Top}(G) = \text{Top}(H)$  if both  $G$  and  $H$  are isomorphic. A topological index is a numerical value related to chemical structure which asserts to show a relationship between chemical structure and various physico-chemical qualities, chemical reactivity, or you could say biological activity [9]. In reality, topological indices are created by converting a molecular graph into a number that describes the topology of that graph. In molecular modeling, we investigate the connection between the structure, characteristics, and activity of chemical substances. Molecular descriptors have a key part in the main chemical structure-related fields like chemistry, pharmacology, etc [10]. In 1998, the generalized Randi index was introduced independently by Bollobas et al. [11] and Amic et al. [12]. Mathematicians and chemists both studied this index [13–17]. Quantitative structure–activity/property/toxicity relationships (QSAR, QSPR, QSTR) modelling is a well-known and well-established field of study. In this type of modelling, physicochemical and molecular descriptors are associated with a bioassay of a medication that is designed to elicit a conventional pharmacological response. These connections have been recognized for a long time as an essential part of the drug discovery and development process. They provide major insights into the role that molecular characteristics play in the biological activity of compounds that are both similar to one another and unrelated to one another. The descriptor-based QSPR techniques make use of the entire structure of the compounds by modelling the property as a function of chemical descriptors. This allows for the most accurate predictions possible. These descriptors might simply be the quantity of such atoms, or they could be features that characterize the linkages between atoms or the form of

TABLE 2 Six physicochemical properties of anti-tuberculosis drugs.

Drug	MM	XLOGP3	C	LOGP	MP	CCS
Isoniazid	137.14	-0.7	120	-0.7	340.5	125.6
Pyrazinamide	123.11	-0.6	115	-0.6	376	122.8
Ethionamide	166.25	1.1	147	0.5	327	135.8
Levofloxacin	361.4	-0.4	634	2.1	225	188.5
Amikacin	585.6	-7.9	819	-8.78	214	235.1
Ofloxacin	361.4	-0.4	634	-0.39	254	188.5
Cycloserine	102.09	-1.5	929	-0.9	155.5	—
4-aminosalicylic acid	153.14	1.3	160	0.89	150.5	—
Ethambutol	204.31	-0.1	109	0.4	171.5	—
Ciprofloxacin	331.34	-1.1	571	0.28	255	185.3
Bedaquiline	555.5	7.2	715	7.74	118	—
Kanamycin	484.5	-6.9	638	-6.3	—	206.6
Streptomycin	581.6	-8	940	-2.53	230	231

TABLE 3 Statistical analysis for the reducible first Zagreb index.

Properties	N	a	b	c	r	r <sup>2</sup>	F	p
<i>For linear model</i>								
MM	13	121.4199	0.3174	—	0.9916	0.9833	647.5515	0.0000
XLOGP3	13	0.4445	-0.002938	—	-0.4068	0.1655	2.1819	0.1677
C	13	254.0263	0.3989	—	0.7074	0.5004	11.0191	0.0068
LOGP	13	0.5154	-0.001852	—	-0.272	0.074	0.8791	0.3686
MP	12	262.3759	-0.04733	—	-0.3421	0.1171	1.3259	0.2763
CCS	9	128.5696	0.07279	—	0.9733	0.9472	125.6425	0.0000
<i>For quadratic model</i>								
MM	13	96.0478	0.4654	-0.0001	0.9981	0.9963	139.5896	0.0000
XLOGP3	13	-0.3322	0.0016	0.0000	0.4353	0.1895	1.1689	0.3498
C	13	226.0558	0.5621	-0.0001	0.7109	0.5055	5.1118	0.0296
LOGP	13	-0.6961	0.0052	0.0000	0.3736	0.1396	0.8111	0.4716
MP	12	266.488	-0.0719	0.0000	0.3446	0.1188	0.6068	0.566
CCS	9	116.9132	0.128	0.0000	0.9947	0.9894	280.729	0.0000
<i>For logarithmic model</i>								
MM	13	-524.7273	144.3139	—	0.9742	0.9491	85.4682	0.0001
XLOGP3	13	5.0647	-1.1031	—	-0.3301	0.1089	1.5637	0.5241
C	13	-465.9574	165.6227	—	0.6346	0.4027	2.5423	0.0026
LOGP	13	2.6789	-0.5673	—	-0.18	0.0324	1.7364	0.7342
MP	12	361.732	-22.0794	—	-0.3417	0.1168	2.2635	0.8317
CCS	9	-31.3612	34.8091	—	0.9861	0.9725	94.6245	0.0003



TABLE 4 Statistical analysis for the reducible second Zagreb index.

Properties	N	a	b	c	r	r <sup>2</sup>	F	p
<i>For linear model</i>								
MM	13	197.8953	0.03537	—	0.8044	0.6471	20.1735	0.0009
XLOGP3	13	-0.9894	-0.0001	—	-0.1163	0.01353	0.1509	0.7051
C	13	347.2226	0.04531	—	0.5849	0.3421	5.7199	0.0357
LOGP	13	-1.114	0.0001	—	0.1487	0.02212	0.2488	0.6278
MP	12	255.5193	-0.0068	—	-0.3499	0.1224	1.3947	0.2649
CCS	9	152.867	0.0075	—	0.6928	0.48	6.4616	0.0385
<i>For quadratic model</i>								
MM	13	158.6936	0.0776	0.0000	0.8410	0.7074	12.0901	0.0021
XLOGP3	13	-0.4226	-0.0007	0.0000	0.1957	0.0383	0.1991	0.8226
C	13	298.5519	0.0977	0.0000	0.6099	0.372	2.9624	0.0376
LOGP	13	-0.407	-0.0006	0.0000	0.2559	0.0655	0.3503	0.7128
MP	12	256.7221	-0.0082	0.0000	0.3502	0.1227	0.6292	0.5549
CCS	9	139.8961	0.0183	0.0000	0.7547	0.5696	3.97	0.0497
<i>For logarithmic model</i>								
MM	13	-356.0103	95.0916	—	0.886	0.7849	63.4267	0.0032
XLOGP3	13	1.6578	-0.4286	—	-0.177	0.0313	0.7357	0.5248
C	13	-208.1414	100.0905	—	0.5293	0.2802	3.7453	0.0005
LOGP	13	-1.2177	0.0817	—	0.0358	0.0013	1.7345	0.7254
MP	12	338.2667	-14.9125	—	-0.3129	0.0979	0.7524	0.0735
CCS	9	10.3761	23.0143	—	0.8559	0.7325	72.5368	0.0001

the molecule in three dimensions. In this manner, a 1D descriptor is a one-dimensional linear representation of the molecule, 2D descriptors are a representation of the molecule in two dimensions on a plane, and 3D descriptors are a representation of the molecule in three dimensions of space. The majority of topological and connectivity indexes can be rearranged using two-dimensional descriptors. Three-dimensional descriptors provide a concise summary of the geometry, the surface, and the volume of the molecule. Quantitative structure-activity relationships (QSARs) are mathematical models created to correlate several types of biological activity, chemical reactivity, equilibrium, physical, and physicochemical qualities [18]. In the developing of QSAR models, twenty primary categories of mistakes have been identified by Dearden et al. They do this by utilizing bad data, overfitting, and bad training and test sets [19]. Simply see Figure 3 for the general overview of the QSPR method. Check out [20–24] for further details on QSPR analysis across various different fields.

The main objective of this article is to develop a quantitative structure-property relationship (QSPR) analysis for tuberculosis (TB) drugs.

- Firstly, we collected data on 13 drugs used in the treatment of TB, along with their six physicochemical properties. These properties were obtained from reputable online sources such as PubChem and ChemSpider.
- Next, we calculated the numerical results of ten degree-based reducible indices to analyze the molecular structure of the 13 TB drugs. Three techniques, namely, edge partition, vertex degree, and counting degree, were used to calculate the ten discussed reducible indices.
- The QSPR analysis was then used to establish a strong positive relationship between the indices and properties. To conduct this analysis, we employed three regression equations: linear-regression, quadratic-regression, and logarithmic-regression. These models were chosen for their strong significance in determining the relationship between the properties and indices.
- One of the statistical parameters used, known as the correlation coefficient ( $r$ ), provides insight into the reliability and significance level between the physical properties and calculated the numerical values of the TB drugs.
- Finally, we utilized line graphs to visually compare all correlation coefficients based on the numerical data. This graphical representation allows for a comprehensive discussion of the relationships between the properties and indices.

In various fields, this paragraph covers the novel progress in the topic of topological indices. In 2023, Abid Mahboob et al. [25] introduced eight novel indices based on the degrees method. They determine the QSPR analysis by utilizing linear, quad, and log models to establish

TABLE 5 Statistical analysis for the reducible reciprocal Randic index.

Properties	N	a	b	c	r	r <sup>2</sup>	F	p
<i>For linear model</i>								
MM	13	120.8357	0.6734	—	0.9921	0.9842	685.6292	0.0000
XLOGP3	13	0.3582	-0.005922	—	-0.3867	0.1495	1.9339	0.1918
C	13	255.4439	0.8391	—	0.7016	0.4923	10.6656	0.0075
LOGP	13	0.4514	-0.003701	—	-0.2563	0.06569	0.7734	0.398
MP	12	262.8292	-0.1017	—	-0.3468	0.1203	1.3674	0.2694
CCS	9	127.921	0.1565	—	0.9751	0.9508	135.2157	0.0000
<i>For quadratic model</i>								
MM	13	96.7815	0.973	-0.0004	0.9980	0.9962	1318.2356	0.0000
XLOGP3	13	0.0001	0.0001	0.0000	0.3987	0.159	0.9451	0.4208
C	13	230.722	1.1469	-0.0004	0.7045	0.4964	4.9279	0.0324
LOGP	13	-0.5473	0.0087	0.0000	0.3339	0.1115	0.6272	0.6272
MP	12	263.1149	-0.1053	0.0000	0.3484	0.1203	0.6154	0.5617
CCS	9	116.6876	0.2695	-0.0002	0.9951	0.9903	305.2349	0.0000
<i>For logarithmic model</i>								
MM	13	-399.3469	141.2089	—	0.9697	0.9402	163.6474	0.0002
XLOGP3	13	3.9097	-1.0407	—	-0.3168	0.1003	1.6376	0.5237
C	13	-283.9748	154.5721	—	0.6025	0.363	6.7263	0.00245
LOGP	13	2.0468	-0.5277	—	-0.1704	0.029	0.6245	0.0856
MP	12	336.1482	-20.3144	—	-0.3202	0.1025	0.7653	0.6254
CCS	9	-6.5383	35.0199	—	0.985	0.9702	39.7256	0.0076

relationships between the novel indices and physical properties of 18 molecular structures of cancer drugs. The significance of all the results is evident as they meet the criteria of a  $p$ -value ( $\leq 0.05$ ) and an  $F$ -test value ( $> 2.5$ ). For example, the reducible 1st Zagreb index proves to be the most accurate predictor for MR, MW, and P, with correlation values reaching 0.9. Abid Mahboob et al. [26] introduced two new terminal descriptors called the “first” and “second” terminal Zagreb indices. These descriptors enhance the QSPR analysis by utilizing linear and quad models to determine the best fit predictor for seven physical properties of 23 molecular structures associated with kidney cancer drugs. The calculations revealed that all correlation “ $r$ ” values exceeded 0.7, providing substantial evidence of a significant relationship between the indices and drug properties. Muhammad Nazri Husin and Anamila Ariffin [27] conducted a comprehensive study on various graph networks, including chain silicate networks  $CS_1$  and  $CS_m$ , hexagonal networks  $HX_2$ ,  $HX_3$ , and  $HX_4$ , and oxide networks  $OX_1$  and  $OX_2$ . They focused on analyzing the generalize the Randic that is based on edge, geometric arithmetic, ABC, MABC, Zagreb indices. The researchers employed the concept of line graphs to investigate these networks. In their study, Jia-Bao Liu et al. [28] conducted calculations to determine various numerical descriptors derived from multiplicative degrees. These descriptors were specifically applied to analyze the structures of three networks: the planar octahedron network [POH(2)], the triangular prism network [TP(2)], and the dominating planar octahedron network [DPOH(2)]. Muhammad Shoaib Sardaret et al. [29] conducted calculations to determine various topological indices using a degree-based technique. These descriptors were calculated to analyze the structures of double silicon carbide  $Si_2C_3-I(p,q)$  where  $p = 1$  and  $q = 1$  and its  $D[Si_2C_3-I(1,1)]$ , as well as the strong double graph of silicon carbide and its  $D[Si_2C_3-I(1,1)]$ . Additionally, they compared the topological indices both numerically and graphically. Zahid Raza et al. [30] calculated the expected values of the first Zagreb connection index in various structures, including a random cyclooctatetraene chain, random polyphenyls chain, and random chain network with  $l$  consisting of octagons, hexagons, and pentagons. They analyzed the maximum and minimum chains with  $COC_b$ ,  $PPC_b$ , and  $PG_b$ , specifically focusing on the expected values of these chains. In 2024, Imran Nadeem et al. [31] conducted a study on the inequalities between general Randic type graph invariants, specifically the general Randic and zeroth-general Randic indices. These indices are calculated based on the vertex-degree methodology. The general Randic index is denoted as  $R_\alpha$ , while the zeroth-general Randic index is denoted as  $Q_\alpha$ . The researchers identified linear inequality relationships between these graph invariants. Additionally, they posed an open problem: to determine the linear inequality between the general Randic index  $R_\alpha$  and the zeroth-general Randic index  $Q_\alpha$  for any negative real number  $\alpha$ .

TABLE 6 Statistical analysis for the reducible reciprocal sum connectivity index.

Properties	N	a	b	c	r	r <sup>2</sup>	F	p
<i>For linear model</i>								
MM	13	81.6943	2.0309	—	0.9981	0.9962	2911.0842	0.0000
XLOGP3	13	0.6249	-0.0172	—	-0.3746	0.1403	1.7954	0.2073
C	13	207.3057	2.525	—	0.7044	0.4962	10.833	0.0071
LOGP	13	0.5319	-0.01001	—	-0.2312	0.05348	0.6215	0.4472
MP	12	268.2718	0.4817	—	-0.3441	0.1184	1.3433	0.2734
CCS	9	116.8003	0.4817	—	0.9923	0.9848	452.0602	0.0000
<i>For quadratic model</i>								
MM	13	75.7318	2.1967	-0.0006	0.9982	0.9966	1448.3066	0.0000
XLOGP3	13	-0.4058	0.0115	-0.0001	0.3991	0.1593	0.1593	0.4199
C	13	231.8508	1.8426	0.0026	0.7056	0.4979	4.9591	0.0319
LOGP	13	-0.9455	0.0311	-0.0002	0.3120	0.0974	0.5398	0.5989
MP	12	264.3534	-0.1936	-0.0004	0.3452	0.1192	0.6088	0.5659
CCS	9	109.0213	0.6722	-0.0007	0.9973	0.9947	567.848	0.0000
<i>For logarithmic model</i>								
MM	13	-79.3429	182.2592	—	0.9679	0.9368	57.5364	0.0062
XLOGP3	13	4.4663	-1.3357	—	-0.3144	0.0989	2.1677	0.8544
C	13	-389.4377	203.5928	—	0.6137	0.3766	28.5367	0.0045
LOGP	13	2.2672	-0.6632	—	-0.1656	0.0274	0.6534	0.0955
MP	12	350.0164	-26.7631	—	-0.3264	0.1066	2.2675	0.6244
CCS	9	-24.2619	44.7407	—	0.9817	0.9637	45.7354	0.0005

## 2 Fundamental definitions with literature review

Let  $G = [V(G), E(G)]$  be a simple graph having  $|V(G)|$  the order and  $|E(G)|$  the size of a graph, where  $V(G)$  is considered as node set and  $E(G) \subseteq V(G) \times V(G)$  is a bond set. Every vertex is considered as an atom in a graph, and bonding within the two atoms is known as edge. The valency or degree of any node is the number of total edges which are incident to the node. Now, few useful reducible TI's is explained which is given below:

### 2.1 Reducible first and second Zagreb indices

Gutman and Trinajstić [32] introduced the two graph invariants known as the first Zagreb index, denoted as  $M_1$ , is a graph theoretic concept used to measure the molecular size or shape of a chemical compound. It takes into account the sum of the degrees of all the vertices in a molecule. On the other hand, the second Zagreb index, denoted as  $M_2$ , calculates the sum of products of degrees for pairs of adjacent vertices in a graph. This index is related to the presence and distribution of bond lengths within a compound. These indices serve as valuable tools in exploring and understanding molecular structures with significant applications across diverse fields including drug discovery, material science, and environmental monitoring. S.R. Islam and M.Pal examined the relationship between the second Zagreb index and physicochemical properties in QSPR research of fuzzy octane isomers graphs [33]. It is their conclusion that this index provides a significantly accurate estimation of the acentric factor and entropy. For instance, the value of acentric factor and entropy for octane isomers are taken from [34]. They calculate the correlation coefficient of  $M_2$  with acentric factor is  $-0.977966531$  and the correlation coefficients of the  $M_2$  with entropy is  $-0.91961647$ . These outcomes declare the appropriateness of this index for octane isomers in chemical compounds. The Zagreb indices and their variations have been utilized to study ideas such as heterosystems, chirality, ZE-isomerism, and molecular complexity.

Abid Mehboob et al. [25] were motivated by the framework of these indices to propose the idea of reducible first and second Zagreb indices. In QSPR modeling and chemical informatics, these indices has produced outstanding outcomes by establishing links between

TABLE 7 Statistical analysis for the reducible first hyper Zagreb index.

Properties	N	a	b	c	r	r <sup>2</sup>	F	p
<i>For linear model</i>								
MM	13	162.7938	0.00754	—	0.9536	0.9094	110.3931	0.0000
XLOGP3	13	0.3402	-0.00008324	—	-0.4666	0.2177	3.0619	0.108
C	13	308.3285	0.009365	—	0.6723	0.452	9.074	0.0118
LOGP	13	0.5904	-0.00005927	—	-0.3524	0.1242	1.5598	0.2376
MP	12	254.006	-0.0009953	—	-0.2934	0.08611	0.9423	0.3546
CCS	9	141.1509	0.001613	—	0.9029	0.8152	30.8694	0.0008
<i>For quadratic model</i>								
MM	13	122.6824	0.0149	0.0000	0.9844	0.9691	156.8696	0.0000
XLOGP3	13	-0.4118	0.0001	0.0000	0.4090	0.259	1.7475	0.2234
C	13	267.8934	0.0168	0.0000	0.6867	0.4716	4.4622	0.0412
LOGP	13	-0.5996	0.0002	0.0000	0.4903	0.2404	1.5823	0.2529
MP	12	259.4173	-0.0021	0.0000	0.3024	0.0915	0.4532	0.6493
CCS	9	124.5373	0.0038	0.0000	0.9549	0.912	31.0833	0.0007
<i>For logarithmic model</i>								
MM	13	-544.4697	96.6785	—	0.9306	0.8661	62.5484	0.0003
XLOGP3	13	5.6758	-0.7905	—	-0.3373	0.1138	1.3476	0.2545
C	13	-348.72	95.2908	—	0.5207	0.2711	37.3654	0.0006
LOGP	13	3.3129	-0.4423	—	-0.2002	0.0401	0.7547	0.3685
MP	12	315.7988	-9.2112	—	-0.2037	0.0415	1.3564	0.0745
CCS	9	-51.3892	24.7484	—	0.9173	0.8414	174.5758	0.0001

the structural characteristics of molecules and their corresponding properties or activities.  $RM_1$  is the notation for this index, which represents the total sum of the “n” degrees of all vertices in a relevant molecular graph, while  $RM_2$  denotes the second reducible zagreb index, defined as the summation of products of the “n” degrees for pairs of adjacent vertices in a molecular graph. In which  $d_u$  and  $d_v$  represent the degrees of the vertices while “n” indicates the order of the vertices. These indices are defined mathematically as:

$$RM_1(G) = \sum_{uv \in E(G)} \left( \frac{n}{d_u} + \frac{n}{d_v} \right),$$

$$RM_2(G) = \sum_{uv \in E(G)} \left( \frac{n}{d_u} \times \frac{n}{d_v} \right).$$

## 2.2 Reducible reciprocal randic index

Some of the most well-known topological indices in regard to chemistry and chemical graph theory was proposed by Milan Randic in 1975 [35] and is useful for evaluating the degree of branching of the carbon atom skeleton of saturated hydrocarbons. The first genuine degree-based topological index in history was the Randic index. This index can help chemists predict certain properties of molecules such as boiling points or toxicity levels. By knowing the level of complexity in a molecule using the randic index, scientists can make informed decisions about its behavior and potential uses. Mingao Yuan derived the limits of the Randic index and its variants for an inhomogeneous Erdos-Rényi random graph [36]. These results shed light on the impact of network heterogeneity on the indices and offer new insights into the Randic index and its variants.

TABLE 8 Statistical analysis for the reducible second hyper Zagreb index.

Propeties	N	a	b	c	r	r <sup>2</sup>	F	p
<i>For linear model</i>								
MM	13	210.4404	0.00007	—	0.7775	0.8818	38.4358	0.0000
XLOGP3	13	-0.02199	-0.0000	—	-0.4904	0.2405	3.4839	0.0888
C	13	366.7098	0.000097	—	0.6253	0.3911	7.0641	0.0222
LOGP	13	0.4716	-0.0000	—	-0.4235	0.1794	2.404	0.1493
MP	12	248.8652	-0.000010	—	-0.2872	0.08246	0.8987	0.3655
CCS	9	155.0311	0.00001	—	0.8217	0.6752	14.5521	0.0065
<i>For quadratic model</i>								
MM	13	178.4115	0.0002	0.0000	0.9548	0.9117	51.6331	0.0000
XLOGP3	13	-0.3579	0	0.0000	0.4191	0.2695	1.845	0.2088
C	13	326.4705	0.0002	0.0000	0.6777	0.4593	4.2476	0.0462
LOGP	13	-0.2598	0	0.0000	0.4780	0.3341	2.5081	0.1315
MP	12	257.5147	0	0.0000	0.3841	0.1476	0.1476	0.4873
CCS	9	145.1945	0	0.0000	0.9003	0.8107	12.8492	0.0068
<i>For logarithmic model</i>								
MM	13	-318.8317	54.8104	—	0.9632	0.9278	78.6536	0.0063
XLOGP3	13	3.4866	-0.4186	—	-0.326	0.1063	1.6478	0.0173
C	13	-202.41	60.5619	—	0.6041	0.3649	28.7367	0.0027
LOGP	13	1.9401	-0.2215	—	-0.183	0.0335	0.3645	0.7254
MP	12	336.6867	-8.9537	—	-0.3603	0.1298	1.4538	0.0846
CCS	9	21.5129	13.0429	—	0.981	0.9623	479.3848	0.0001

In order to calculate the reducible randic index, you need to add up the square root of the distance between each bond. This distance is calculated by multiplying the “n” degrees of the vertices in a molecular graph. This index are defined mathematically as:

$$RR(G) = \sum_{uv \in E(G)} \sqrt{\frac{n}{d_u} \times \frac{n}{d_v}}$$

## 2.3 Reducible reciprocal sum connectivity index

Zhou and Trinajstic [37] were the ones who initially brought the idea of the sum connectivity index. This index was developed by drawing inspiration from the Randic-index. It is a concept used in chemistry to measure how connected different atoms are within a molecule. It is a way to determine the overall structural complexity of a molecule and its ability to interact with other molecules. This index is takes into account factors like bond lengths, angles, and types of bonds between atoms to calculate a numerical value. Muhammad Hussain et al. [38] have conducted a comprehensive analysis to determine the sum connectivity index for three subdivisions of the  $H_3BO_3$  layer structure. These subdivisions include the subdivision  $S(H_3BO_3)$ , the line graph  $L(H_3BO_3)$ , and the caged chain  $C(H_3BO_3)$ . The researchers employed various mathematical and computational methods to accurately calculate the sum connectivity index for each subdivision. The correlation between this index and the straight line graph is significant [39].

A novel index known as the reducible sum connectivity index has been introduced. It is calculated by summing the square root of each bonds distance, where the distance is obtained by summing the “n” degrees of the connecting vertices, and has a mathematical definition as:

$$RSC(G) = \sum_{uv \in E(G)} \sqrt{\frac{n}{d_u} + \frac{n}{d_v}}$$

TABLE 9 Statistical analysis for the reducible sigma index.

Properties	N	a	b	c	r	r <sup>2</sup>	F	p
<i>For linear model</i>								
MM	13	192.419	0.04944	—	0.8966	0.8039	45.0836	0.0000
XLOGP3	13	0.7616	-0.0008381	—	-0.6737	0.4538	9.1398	0.0115
C	13	335.0517	0.06534	—	0.6726	0.4524	9.0879	0.0117
LOGP	13	0.9285	-0.0006116	—	-0.5214	0.2719	4.1076	0.0676
MP	12	246.1553	-0.004883	—	-0.2049	0.042	0.4384	0.5229
CCS	9	146.93	0.01019	—	0.8899	0.792	26.6509	0.0013
<i>For quadratic model</i>								
MM	13	138.522	0.129	0.0000	0.9766	0.9538	103.3206	0.0000
XLOGP3	13	-0.2409	0.0006	0.0000	0.7455	0.5558	6.2551	0.0173
C	13	280.8399	0.1454	0.0000	0.7080	0.5013	5.0259	0.0308
LOGP	13	0.0229	0.0007	0.0000	0.6045	0.3655	2.8798	0.1029
MP	12	267.6473	-0.04	0.0000	0.3945	0.1557	0.8299	0.4669
CCS	9	133.0309	0.0276	0.0000	0.9550	0.9121	31.1128	0.0007
<i>For logarithmic model</i>								
MM	13	-294.8106	93.1536	—	0.9729	0.9465	368.6357	0.0002
XLOGP3	13	4.4765	-0.8895	—	-0.4117	0.1695	1.3865	0.8384
C	13	-206.2213	107.5349	—	0.6375	0.406	4.5786	0.0028
LOGP	13	2.8294	-0.5261	—	-0.2583	0.0667	0.3746	0.0628
MP	12	322.1681	-13.6067	—	-0.3225	0.104	1.3654	0.2786
CCS	9	27.9428	21.7919	—	0.9922	0.9844	48.6577	0.0006

## 2.4 Reducible 1st and 2nd hyper Zagreb indices

Shirdel et al. [40] presented the concept of the hyper Zagreb index, which is a molecular descriptor referred to as distance-based Zagreb index. These indices are numerical values that provide important information about a molecule's structure. The first hyper Zagreb index is obtained by summing up the squares of the degrees of all vertices in a molecule. On the other hand, the second hyper Zagreb index is calculated by multiplying the degrees of adjacent vertices and then summing them all up. By examining both indices, researchers can gain insight into the complexity of a molecule and make predictions about its chemical properties. Ishita Sarkar and Manjunath Nanjappa [41] conducted a study on the combinatorial inequalities related to edges, vertices, and their corresponding neighborhood notions. They also incorporated other molecular descriptors in their calculations to determine the exact expressions of the second hyper-Zagreb index for specific corona products involving the semi-total point graph. Abdul Hakeem et al. [42] studied the QSPR analysis on heart attack disease using degree-based indices. The researchers found that the first hyper Zagreb index exhibited a robust correlation with molecular weight and boiling point within the range of 0.9. Similarly, the second hyper Zagreb index demonstrated a strong correlation with boiling point within the same range.

In the molecular graph, the reducible first hyper Zagreb index is obtained by summing up the squares of the “n” degrees of all vertices. On the other hand, the reducible second hyper Zagreb index is calculated by multiplying the “n” degrees of adjacent vertices and then summing them all up. Mathematically these indices is defined as:

$$RHM_1(G) = \sum_{uv \in E(G)} \left( \frac{n}{d_u} + \frac{n}{d_v} \right)^2,$$

$$RHM_2(G) = \sum_{uv \in E(G)} \left( \frac{n}{d_u} \times \frac{n}{d_v} \right)^2.$$

TABLE 10 Statistical analysis for the reducible forgotten index.

Properties	N	a	b	c	r	r <sup>2</sup>	F	p
<i>For linear model</i>								
MM	13	169.4351	0.01312	—	0.9571	0.9161	120.0949	0.0000
XLOGP3	13	0.3495	-0.0001521	—	-0.4918	0.2419	3.5092	0.0878
C	13	312.626	0.01665	—	0.6892	0.4749	9.9497	0.0091
LOGP	13	0.5932	-0.000108	—	-0.3703	0.1371	1.7477	0.213
MP	12	253.8904	-0.001809	—	-0.3071	0.09431	1.0412	0.3316
CCS	9	143.0493	0.002774	—	0.9169	0.8407	36.9467	0.0005
<i>For quadratic model</i>								
MM	13	132.53	0.0267	0.0000	0.9934	0.987	379.2814	0.0000
XLOGP3	13	-0.4879	0.0002	0.0000	0.456	0.3136	2.2842	0.1524
C	13	268.6353	0.0329	0.0000	0.7123	0.5074	5.15	0.0296
LOGP	13	-0.5548	0.0003	0.0000	0.4373	0.2887	2.0298	0.1826
MP	12	266.2883	-0.0068	0.0000	0.3672	0.1349	0.7015	0.5211
CCS	9	128.9797	0.0067	0.0000	0.9757	0.9521	59.6012	0.0001
<i>For logarithmic model</i>								
MM	13	465.5766	95.8217	—	0.9742	0.949	63.3654	0.0027
XLOGP3	13	4.9133	-0.7692	—	-0.3466	0.1201	1.3745	0.7253
C	13	-357.8937	105.0634	—	0.6063	0.3676	73.7354	0.0235
LOGP	13	2.7509	-0.4138	—	-0.1978	0.0391	0.3876	0.0975
MP	12	347.2109	-13.9873	—	-0.3252	0.1058	2.1836	0.0754
CCS	9	-19.4979	23.3651	—	0.9901	0.9803	94.7648	0.0001

## 2.5 Reducible sigma index

The concept of Albertson index served as motivation for Gutman, leading him to propose the idea of the Sigma index [43]. His paper explores the inverse problem concerning the sigma index, demonstrating that an even value will always be obtained for this index in any given graph. Merve and Ismail conducting a calculation of the sigma index for various graphs, adding the subdivision of the cycle graph  $C_m$ , the graph of  $S(G)$ , and the  $r$ -subdivision graph of  $S'(G)$  [44]. Additionally, they provide examples of well-known graph cycles. Ozge Colakoglu Havare [45] studied the QSPR analysis with curvilinear regression model by using various molecular descriptors including Sigma index for the monocarboxylic acids.

The reducible sigma index can be obtained by subtracting the squares of the “ $n$ ” degrees of each vertex in a molecule. This index is mathematically defined as:

$$RS(G) = \sum_{uv \in E(G)} \left( \frac{n}{d_u} - \frac{n}{d_v} \right)^2.$$

## 2.6 Reducible forgotten index

Furtula and Gutman conducted an analysis on the latest version of Zagreb indices, referred to as the Forgotten index [46]. Its definition lies in using the sum of cubes of degrees in molecular graphs to examine the influence of structure dependency on different properties. The purpose of presenting this index is to demonstrate its numerous significant features and emphasize its exceptional effectiveness in enhancing the chemical attributes of the original Zagreb index. This index plays a crucial role in significantly enhancing the physico-chemical applicability of the first Zagreb index. Some upper bounds for the F-index and certain applications were computed by Hosamani [47]. The precise formulas for the graph operations with regards to the F-index were calculated by Nilanjan and other authors [48]. Recently, the reducible forgotten index is introduced which is mathematically defined as:

TABLE 11 Statistical analysis for the reducible first Gourava index.

Properties	N	a	b	c	r	r <sup>2</sup>	F	p
<i>For linear model</i>								
MM	13	160.963	0.03131	—	0.9667	0.9344	156.8033	0.0000
XLOGP3	13	0.1775	-0.0003094	—	-0.4234	0.1793	2.4032	0.1494
C	13	304.648	0.03917	—	0.4712	0.6864	9.8003	0.0095
LOGP	13	0.4452	-0.0002145	—	-0.3114	0.09694	1.1809	0.3004
MP	12	256.6838	-0.004654	—	-0.335	0.1122	1.2637	0.2872
CCS	9	140.455	0.006874	—	0.9283	0.8618	43.6559	0.0003
<i>For quadratic model</i>								
MM	13	124.4735	0.016	0.0000	0.9959	0.9919	614.9352	0.0000
XLOGP3	13	-0.253	0	0.0000	0.4415	0.195	1.2113	0.3387
C	13	256.2399	0.0785	0.0000	0.7097	0.5038	5.0758	0.0301
LOGP	13	-0.4838	0.0005	0.0000	0.4234	0.1793	1.0925	0.3723
MP	12	263.8316	-0.0109	0.0000	0.3511	0.1233	0.6329	0.5531
CCS	9	126.0261	0.0155	0.0000	0.9834	0.9672	88.5105	0.0000
<i>For logarithmic model</i>								
MM	13	-463.4443	104.0763	—	0.974	0.9486	174.7654	0.0006
XLOGP3	13	4.5578	-0.7904	—	-0.3279	0.1075	1.3755	0.8548
C	13	-346.22	112.8723	—	0.5996	0.3595	37.7254	0.0075
LOGP	13	2.4916	-0.4162	—	-0.1831	0.0335	1.7542	0.0645
MP	12	343.4525	-14.7187	—	-0.3157	0.0997	0.6432	0.8654
CCS	9	-23.0011	25.8757	—	0.9896	0.9793	275.8643	0.0076

$$RF(G) = \sum_{uv \in E(G)} \left( \left( \frac{n}{d_u} \right)^2 + \left( \frac{n}{d_v} \right)^2 \right).$$

### 2.7 Reducible 1st and 2nd Gourava index

The first and second Gourava indices were introduced by V.R. Kulli [49], who drew inspiration from the definition of the Zagreb indices and their extensive usage. In calculating this index, Kulli applies the equation’s definition to both armchair and zigzag edge polyhex nanotubes, with a particular focus on a few well-known graph types. Ying Wang et al. [50] prove the graphs uniqueness on trees with given independence numbers with maximum both 1st and 2nd Gourava indices. The mathematical definition of the reducible first and second Gourava indices was recently introduced, which is stated as:

$$RG_1(G) = \sum_{uv \in E(G)} \left( \frac{n}{d_u} + \frac{n}{d_v} \left( \frac{n}{d_u} + \frac{n}{d_v} \right) \right),$$

$$RG_2(G) = \sum_{uv \in E(G)} \left( \left( \frac{n}{d_u} + \frac{n}{d_v} \right) \left( \frac{n}{d_u} \times \frac{n}{d_v} \right) \right).$$

## 3 Methods and strategies

Topological indices are derived from various mathematical approaches, and one of the most popular methodologies is based on vertex degrees. The degree of a vertex in a graph represents the number of edges connected to it. By analyzing the distribution of degrees in a system, we can obtain



TABLE 12 Statistical analysis for the reducible second Gourava index.

Properties	N	a	b	c	r	r <sup>2</sup>	F	p
<i>For linear model</i>								
MM	13	191.5845	0.0007926	—	0.9252	0.856	65.3868	0.0000
XLOGP3	13	0.09156	-0.00000918	—	0.2256	-0.475	3.2051	0.1009
C	13	342.2894	0.0009956	—	0.6597	0.4352	8.4765	0.0141
LOGP	13	0.4739	-0.000006913	—	-0.3794	0.1439	1.8496	0.2011
MP	12	251.5699	-0.0001121	—	-0.307	0.09427	1.0408	0.3317
CCS	9	149.7374	0.0001628	—	0.8707	0.7581	21.9342	0.0022
<i>For quadratic model</i>								
MM	13	156.491	0.0018	0.0000	0.9794	0.9594	118.2158	0.0000
XLOGP3	13	-0.3882	0	0.0000	0.4134	0.2636	1.7899	0.2165
C	13	297.8582	0.0023	0.0000	0.6989	0.4886	4.7777	0.0356
LOGP	13	-0.4004	0	0.00000	0.5346	0.2858	2.0013	0.1858
MP	12	262.0989	-0.0005	0.0000	0.3806	0.1449	0.7626	0.4944
CCS	9	137.149	0.0004	0.0000	0.9445	0.8921	24.8128	0.0013
<i>For logarithmic model</i>								
MM	13	-433.6237	73.2963	—	0.9727	0.9461	356.6357	0.0087
XLOGP3	13	4.4809	-0.5712	—	-0.336	0.1129	1.3846	0.6432
C	13	-298.8449	78.0269	—	0.5877	0.3454	34.6535	0.0036
LOGP	13	2.5081	-0.3063	—	-0.1911	0.0365	0.6532	0.6543
MP	12	338.4152	-10.2879	—	-0.3124	0.0976	1.7363	0.7254
CCS	9	-16.0368	18.2656	—	0.9906	0.9813	476.7256	0.0007

valuable information about its structural features and properties. Degree-based topological indices are calculated using mathematical formulas that take into account the degrees of vertices in a graph. We took several steps to get at the results, including gathering data on the pharmaceutical properties of medications used to treat tuberculosis disease. The data for these properties has been collected from the PubChem website. The numerical values of the molecular descriptors are determined by employing the techniques of vertex degree and edge division. This method involves grouping or bundling the edges and vertices based on their specific attributes. To develop the QSPR analysis, a three regression-equation (linear-regression, quadratic-regression, and logarithmic-regression) and eight statistical parameters (including *p*-value and *F*-test value) were employed. Analyzing and predicting data patterns is made easier with the use of these valuable models. The coefficient correlation *r* results of these three regression-model are determined through the use of the SPSS software. The illustration of the 2D molecular structures of these medications is achieved through the use of software applications like Chem-Draw and Chem-Doodle. By visually representing the intricate structures, these software enable a deeper understanding of the chemical properties of various compounds. Bar graphs have been employed to visually represent the comparisons of all correlation values. These graphs are created through Microsoft Excel. All indices can be calculated using software like Matlab, whereas Maple has been exclusively designed for graph sketching purposes. This methodology has the advantage that all the indices can be computed rapidly. This manuscript presents a concise and comprehensible flowchart of the work, using Figure 4 to make it easy and understandable for readers.

## 4 Analyzing the structural and mathematical computations associated with anti-tuberculosis drugs

This section provides an overview of the structural information and computational analysis of reducible indices for the treatment of tuberculosis (TB) disease. The article primarily focuses on the planar and 2D forms of drugs, disregarding their actual 3D structures. Within this context, the discussion revolves around thirteen prominent anti-tuberculosis drugs, namely, Isoniazid, Pyrazinamide, Ethionamide, Levofloxacin, Amikacin, Ofloxacin, Cycloserine, 4-aminosalicylic acid, Ethambutol, Ciprofloxacin, Bedaquiline,

Kanamycin, and Streptomycin. The thirteen anti-tuberculosis drugs mentioned have been extensively studied and utilized in the treatment of TB. Each drug possesses unique characteristics and mechanisms of action, making them valuable tools in combating this infectious disease. Figure 5 visually displays the chemical structures of these medicine, which are commonly used to treat tuberculosis disease. Additionally, this section provides a comprehensive overview of both the experimental data and the calculation of reducible indices. The six physicochemical properties: molar mass, XLOGP3, Complexity, LOGP, melting point and collision cross section were obtained from the PubChem website and are presented in Table 2. The values of the ten degree-based reducible topological indices were computed using the formulas illustrated in Table 1. All the indices considered in this study are based on vertex and edge degrees. It is noting that a total of thirteen structures were analyzed, each exhibiting eight distinct types of edge bundles. In addition, the maximum vertex degree is restricted to four, and thus the node degree varies between one and four. Therefore, these eight edge bundles are:

$$\begin{aligned}
 E_{(1,2)} &= e = uv \in E(G) | d_u = 1, d_v = 2, E_{(2,2)} = e = uv \in E(G) | d_u = 2, d_v = 2 \\
 E_{(1,3)} &= e = uv \in E(G) | d_u = 1, d_v = 3, E_{(3,3)} = e = uv \in E(G) | d_u = 3, d_v = 3 \\
 E_{(1,4)} &= e = uv \in E(G) | d_u = 1, d_v = 4, E_{(2,4)} = e = uv \in E(G) | d_u = 2, d_v = 4 \\
 E_{(2,3)} &= e = uv \in E(G) | d_u = 2, d_v = 3, E_{(3,4)} = e = uv \in E(G) | d_u = 3, d_v = 4
 \end{aligned}$$

The use of isoniazid, along with other medications, is necessary to treat active tuberculosis (TB) infections. Additionally, when an individual tests positive for TB through a skin test and is suspected of having the infection, isoniazid is administered alone to prevent the development of active TB infections. As an antibiotic, isoniazid functions by preventing bacterial growth. Only bacterial illnesses are treated by this antibiotic. Conversely, it proves ineffective against viral illnesses such as the flu or the common cold. The molecular formula of isoniazid is  $C_6H_7N_3O$ . It consists of carbon (C), hydrogen (H), nitrogen (N), and oxygen (O) atoms. Figure 6 illustrates the 2D molecular structure of Isoniazid. The molecular structure of Isoniazid, denoted as G, is composed of 10 vertices (dots) and 10 edges (lines). To analyze this structure, a technique of degree-based has been employed. The vertex set of this structure,  $V(Isoniazid)$ , consists of vertices  $V_v$ , where v ranges from 1 to 10. Similarly, the edge set,  $E(Isoniazid)$ , is formed by the edges  $\{v_1v_2, v_2v_3, v_3v_4, \dots, v_9v_{10}\}$ . There are three vertex partitions, namely,  $V_1, V_2$ , and  $V_3$ , that can be identified within this structure. In the given graph G,  $V_1$  represents the set of vertices that have a degree of 1. These vertices are connected to only one other vertex in the graph. On the other hand  $V_2$ , consists of vertices that have a degree of 2, meaning they are connected to two other vertices in the graph. Lastly,  $V_3$  represents the set of vertices that have a degree of 3, indicating that they are connected to three other vertices in the graph. Therefore, five edge bundles of this structure is obtained. The cardinalities of these bundles are as follows:  $|E_{(1,2)}| = 1, |E_{(1,3)}| = 1, |E_{(2,3)}| = 3, |E_{(3,3)}| = 1$ , and  $|E_{(2,2)}| = 4$ . These edge bundles are utilized to compute all ten defined reducible indices of the isoniazid structure, which will be discussed below. The same methodology can be applied to determine the remaining indices of other drug structures.

$$\begin{aligned}
 RM_1(Isoniazid) &= \sum_{uv \in E(Isoniazid)} \left( \frac{n}{d_u} + \frac{n}{d_v} \right) = 1 \left( \frac{10}{1} + \frac{10}{2} \right) + 1 \left( \frac{10}{1} + \frac{10}{3} \right) + 3 \left( \frac{10}{2} + \frac{10}{3} \right) + 1 \left( \frac{10}{3} + \frac{10}{3} \right) + 4 \left( \frac{10}{2} + \frac{10}{2} \right) \\
 &= 1(15) + 1(13.3333) + 3(8.3333) + 1(6.6667) + 4(10) = 100
 \end{aligned}$$

$$\begin{aligned}
 RM_2(Isoniazid) &= \sum_{uv \in E(Isoniazid)} \left( \frac{n}{d_u} \times \frac{n}{d_v} \right) = 1 \left( \frac{10}{1} \times \frac{10}{2} \right) + 1 \left( \frac{10}{1} \times \frac{10}{3} \right) + 3 \left( \frac{10}{2} \times \frac{10}{3} \right) + 1 \left( \frac{10}{3} \times \frac{10}{3} \right) + 4 \left( \frac{10}{2} \times \frac{10}{2} \right) \\
 &= 1(50) + 1(33.3333) + 3(16.6667) + 1(11.1111) + 4(25) = 244.4445
 \end{aligned}$$

$$\begin{aligned}
 RR(Isoniazid) &= \sum_{uv \in E(Isoniazid)} \left( \sqrt{\frac{n}{d_u} \times \frac{n}{d_v}} \right) = 1 \left( \sqrt{\frac{10}{1} \times \frac{10}{2}} \right) + 1 \left( \sqrt{\frac{10}{1} \times \frac{10}{3}} \right) + 3 \left( \sqrt{\frac{10}{2} \times \frac{10}{3}} \right) + 1 \left( \sqrt{\frac{10}{3} \times \frac{10}{3}} \right) + 1 \left( \sqrt{\frac{10}{2} \times \frac{10}{2}} \right) \\
 &= 1(15.8113) + 1(10.5409) + 3(5.2704) + 1(3.5137) + 4(7.9057) = 77.2999
 \end{aligned}$$

$$\begin{aligned}
 RSC(Isoniazid) &= \sum_{uv \in E(Isoniazid)} \left( \sqrt{\frac{n}{d_u} + \frac{n}{d_v}} \right) = 1 \left( \sqrt{\frac{10}{1} + \frac{10}{2}} \right) + 1 \left( \sqrt{\frac{10}{1} + \frac{10}{3}} \right) + 3 \left( \sqrt{\frac{10}{2} + \frac{10}{3}} \right) + 1 \left( \sqrt{\frac{10}{3} + \frac{10}{3}} \right) + 1 \left( \sqrt{\frac{10}{2} + \frac{10}{2}} \right) \\
 &= 1(8.1622) + 1(6.4957) + 3(4.9144) + 1(4.3874) + 4(6.5811) = 60.1129
 \end{aligned}$$

$$\begin{aligned}
 RHM_1(Isoniazid) &= \sum_{tw \in E(Isoniazid)} \left( \frac{n}{d_t} + \frac{n}{d_w} \right)^2 = 1 \left( \frac{10}{1} + \frac{10}{2} \right)^2 + 1 \left( \frac{10}{1} + \frac{10}{3} \right)^2 + 3 \left( \frac{10}{2} + \frac{10}{3} \right)^2 + 1 \left( \frac{10}{3} + \frac{10}{3} \right)^2 + 4 \left( \frac{10}{2} + \frac{10}{2} \right)^2 \\
 &= 1(225) + 1(177.7778) + 3(69.4444) + 1(44.4444) + 4(100) = 1055.5555
 \end{aligned}$$

$$\begin{aligned}
 RHM_2(Isoniazid) &= \sum_{uv \in E(Isoniazid)} \left( \frac{n}{d_u} \times \frac{n}{d_v} \right)^2 = 1 \left( \frac{10}{1} \times \frac{10}{2} \right)^2 + 1 \left( \frac{10}{1} \times \frac{10}{3} \right)^2 + 3 \left( \frac{10}{2} \times \frac{10}{3} \right)^2 + 1 \left( \frac{10}{3} \times \frac{10}{3} \right)^2 + 4 \left( \frac{10}{2} \times \frac{10}{2} \right)^2 \\
 &= 1(2500) + 1(1111.1111) + 3(277.7778) + 1(123.4567) + 4(625) = 7067.9012
 \end{aligned}$$

TABLE 13 Comparison table of correlation coefficient of linear regression.

Indices	MM	XLOGP3	C	LOGP	MP	CCS
RM <sub>1</sub> (G)	0.9916	-0.4068	0.7074	-0.272	-0.3421	0.9733
RM <sub>2</sub> (G)	0.8044	-0.1163	0.5849	0.1487	-0.3499	0.6928
RR(G)	0.9921	-0.3867	0.70167	-0.2563	-0.3468	0.9751
RSC(G)	0.9981	-0.3746	0.7044	-0.2312	-0.3441	0.9923
RHM <sub>1</sub> (G)	0.9536	-0.4666	0.6723	-0.3524	-0.2934	0.9029
RHM <sub>2</sub> (G)	0.7775	-0.4904	0.6253	-0.4235	-0.2872	0.8217
RS(G)	0.8966	-0.6737	0.6726	-0.5214	-0.2049	0.8899
RF(G)	0.9571	-0.4918	0.6892	-0.3703	-0.3071	0.9169
RG <sub>1</sub> (G)	0.9667	-0.4234	0.4712	-0.3114	-0.335	0.9283
RG <sub>2</sub> (G)	0.9252	0.2256	0.6597	-0.3794	-0.307	0.8707

TABLE 14 Comparison table of correlation coefficient of quadratic regression.

Indices	MM	XLOGP3	C	LOGP	MP	CCS
RM <sub>1</sub> (G)	0.9981	0.4353	0.7109	0.3736	0.3446	0.9947
RM <sub>2</sub> (G)	0.8410	0.1957	0.6099	0.2559	0.3502	0.7547
RR(G)	0.9980	0.3987	0.7045	0.3339	0.3484	0.9951
RSC(G)	0.9982	0.3991	0.7056	0.3120	0.3452	0.9973
RHM <sub>1</sub> (G)	0.9844	0.4090	0.6867	0.4903	0.3024	0.9549
RHM <sub>2</sub> (G)	0.9548	0.4191	0.6777	0.4780	0.3841	0.9003
RS(G)	0.9766	0.7455	0.7080	0.6045	0.3945	0.9550
RF(G)	0.9934	0.456	0.7123	0.4373	0.3672	0.9757
RG <sub>1</sub> (G)	0.9959	0.4415	0.7097	0.4234	0.3511	0.9834
RG <sub>2</sub> (G)	0.9794	0.4134	0.6989	0.5346	0.3806	0.9445

$$RS(Isoniazid) = \sum_{uv \in E(Isoniazid)} \left( \frac{n}{d_u} - \frac{n}{d_v} \right)^2 = 1 \left( \frac{10}{1} - \frac{10}{2} \right)^2 + 1 \left( \frac{10}{1} - \frac{10}{3} \right)^2 + 3 \left( \frac{10}{2} - \frac{10}{3} \right)^2 + 1 \left( \frac{10}{3} - \frac{10}{3} \right)^2 + 4 \left( \frac{10}{2} - \frac{10}{2} \right)^2$$

$$= 1(25) + 1(44.4444) + 3(2.7778) + 1(0) + 4(0) = 77.7778$$

$$RF(Isoniazid) = \sum_{uv \in E(Isoniazid)} \left( \left( \frac{n}{d_u} \right)^2 + \left( \frac{n}{d_v} \right)^2 \right)$$

$$= 1 \left( \left( \frac{10}{1} \right)^2 + \left( \frac{10}{2} \right)^2 \right) + 1 \left( \left( \frac{10}{1} \right)^2 + \left( \frac{10}{3} \right)^2 \right) + 3 \left( \left( \frac{10}{2} \right)^2 + \left( \frac{10}{3} \right)^2 \right) + 1 \left( \left( \frac{10}{3} \right)^2 + \left( \frac{10}{3} \right)^2 \right) + 4 \left( \left( \frac{10}{2} \right)^2 + \left( \frac{10}{2} \right)^2 \right)$$

$$= 1(125) + 1(111.1111) + 3(36.1111) + 1(22.2222) + 4(50) = 566.6666$$

$$RG_1(Isoniazid) = \sum_{uv \in E(Isoniazid)} \left( \frac{n}{d_u} + \frac{n}{d_v} + \left( \frac{n}{d_u} \times \frac{n}{d_v} \right) \right)$$

$$= 1 \left( \frac{10}{1} + \frac{10}{2} + \left( \frac{10}{1} \times \frac{10}{2} \right) \right) + 1 \left( \frac{10}{1} + \frac{10}{3} + \left( \frac{10}{1} \times \frac{10}{3} \right) \right) + 3 \left( \frac{10}{2} + \frac{10}{3} + \left( \frac{10}{2} \times \frac{10}{3} \right) \right) + 1 \left( \frac{10}{3} + \frac{10}{3} + \left( \frac{10}{3} \times \frac{10}{3} \right) \right)$$

$$+ 4 \left( \frac{10}{2} + \frac{10}{2} + \left( \frac{10}{2} \times \frac{10}{2} \right) \right)$$

$$= 1(65) + 1(46.6667) + 3(25) + 1(17.7778) + 4(35) = 344.4445$$

$$\begin{aligned}
 RG_2(Isoniazid) &= \sum_{uv \in E(Isoniazid)} \left( \left( \frac{n}{d_u} + \frac{n}{d_v} \right) \left( \frac{n}{d_u} \times \frac{n}{d_v} \right) \right) \\
 &= 1 \left( \left( \frac{10}{1} + \frac{10}{2} \right) \left( \frac{10}{1} \times \frac{10}{2} \right) \right) + 1 \left( \left( \frac{10}{1} + \frac{10}{3} \right) \left( \frac{10}{1} \times \frac{10}{3} \right) \right) + 3 \left( \left( \frac{10}{2} + \frac{10}{3} \right) \left( \frac{10}{2} \times \frac{10}{3} \right) \right) + 1 \left( \left( \frac{10}{3} + \frac{10}{3} \right) \left( \frac{10}{3} \times \frac{10}{3} \right) \right) \\
 &\quad + 4 \left( \left( \frac{10}{2} + \frac{10}{2} \right) \left( \frac{10}{2} \times \frac{10}{2} \right) \right) \\
 &= 1(750) + 1(444.4444) + 3(138.8889) + 1(74.0740) + 4(250) = 2685.1851
 \end{aligned}$$

## 5 Regression models

Regression models are fundamental tools in statistics that allow us to understand and analyze the relationship between variables. By fitting a regression model to a dataset, we can make predictions, uncover patterns, and gain insights into the underlying factors that influence a particular outcome. Regression analysis is like playing matchmaker for numbers. It helps us understand the relationship between different variables and predicts how they interact with each other. Basically, it is a statistical method to find patterns and connections within data. Regression models come in different types, Some popular types include linear regression, non-linear regression, and multiple regression analysis. This article covering three regression models such as linear-regression, quadratic-regression and logarithmic-regression.

Linear regression is like a good old-fashioned straight line. It is the simplest type of regression where we try to fit a line that best represents the relationship between an independent variable and a dependent variable. The variable used for prediction is referred to as the dependent variable (Y), while the term used to predict the value of another variable is known as the independent variable (TI). The linear regression equation is in the form:

$$Y = a + b(TI) \quad (5.1)$$

Quadratic regression is a statistical modeling technique that allows us to analyze the relationship between a dependent variable and one or more independent variables. It is an extension of linear regression, where the relationship between variables is assumed to be quadratic rather than linear. Quadratic regression allows us to model relationships with a curved pattern, capturing more complexity than linear regression. The quadratic regression equation is a fancy way of expressing a curved relationship between variables. In quadratic equation (Y) is the dependent variable, (TI) is the independent variable, and a, b, and c are coefficients that determine the shape of the curve. The quadratic regression model comprises three important components: the intercept (a), the linear coefficient (b), and the quadratic coefficient (c). These components determine the position, slope, and curvature of the curve, respectively. The quadratic regression equation is in the form:

$$Y = a + b(TI) + c(TI)^2 \quad (5.2)$$

One commonly used type of regression model is logarithmic regression. Logarithmic regression offers a unique approach to exploring a relationships by fitting a logarithmic function to the data. Logarithmic regression is a type of regression analysis where the relationship between the dependent variable (Y) and the independent variable (TI) is modeled using a logarithmic function. Unlike linear regression, which assumes a linear relationship between the variables, logarithmic regression acknowledges that the relationship may not be linear and allows for a more flexible curve. The logarithmic regression equation is in the form:

$$Y = a + b \cdot \ln(TI) \quad (5.3)$$

The ten reducible indices mentioned above serve as models for six physical properties: molar mass, XLOGP3, Complexity, LOGP, melting point, and collision cross section of the thirteen molecular structures of the anti-tuberculosis drugs. Eqs 5.1–5.3, presented above, are employed to determine the regression equations for degree-based reducible TIs. These equations represent the linear model, quadratic model, and logarithmic model. All three models have been derived using SPSS statistical software. These models are widely used in various fields, including economics, finance, marketing, and social sciences. They provide a quantitative approach to understand the relationships between variables and make predictions based on the available data. By examining the patterns and trends in the data, these regression models help us gain valuable insights and make informed decisions.

### 5.1 Reducible first Zag-index $RM_1(G)$

$MM = 121.4199 + 0.3174(RM_1(G))$	$MM = -0.0001(RM_1(G))^2 + 0.4654(RM_1(G)) + 96.0478$
$XLOGP3 = 0.4445 - 0.002938(RM_1(G))$	$XLOGP3 = 0.0000(RM_1(G))^2 + 0.0016(RM_1(G)) - 0.3322$
$C = 254.0263 + 0.3989(RM_1(G))$	$C = -0.0001(RM_1(G))^2 + 0.5621(RM_1(G)) + 226.0558$
$LOGP = 0.5154 - 0.001852(RM_1(G))$	$LOGP = 0.0000(RM_1(G))^2 + 0.0052(RM_1(G)) - 0.6961$
$MP = 262.3759 - 0.04733(RM_1(G))$	$MP = 0.0000(RM_1(G))^2 - 0.0719(RM_1(G)) + 266.488$
$CCS = 128.5696 + 0.07279(RM_1(G))$	$CCS = 0.0000(RM_1(G))^2 + 0.128(RM_1(G)) + 116.9132$

$MM = -524.7273 + 144.3139.ln(RM_1(G))$   
 $XLOGP3 = 5.0647 - 1.1031.ln(RM_1(G))$   
 $C = -465.9574 + 165.6227.ln(RM_1(G))$   
 $LOGP = 2.6789 - 0.5673.ln(RM_1(G))$   
 $MP = 361.732 - 22.0794.ln(RM_1(G))$   
 $CCS = -31.3612 + 34.8091.ln(RM_1(G))$

### 5.2 Reducible second Zag-index $RM_2(G)$

$MM = 197.8953 + 0.03537(RM_2(G))$	$MM = 0.0000(RM_2(G))^2 + 0.0776(RM_2(G)) + 158.6936$
$XLOGP3 = -0.9894 - 0.0001154(RM_2(G))$	$XLOGP3 = 0.0000(RM_2(G))^2 - 0.0007(RM_2(G)) - 0.4226$
$C = 347.2226 + 0.04531(RM_2(G))$	$C = 0.0000(RM_2(G))^2 + 0.0977(RM_2(G)) + 298.5519$
$LOGP = -1.114 + 0.0001391(RM_2(G))$	$LOGP = 0.0000(RM_2(G))^2 - 0.0006(RM_2(G)) - 0.407$
$MP = 255.5193 - 0.006822(RM_2(G))$	$MP = 0.0000(RM_2(G))^2 - 0.0082(RM_2(G)) + 256.7221$
$CCS = 152.867 + 0.00754(RM_2(G))$	$CCS = 0.0000(RM_2(G))^2 + 0.0183(RM_2(G)) + 139.8961$

$MM = -356.0103 + 95.0916.ln(RM_2(G))$   
 $XLOGP3 = 1.6578 - 0.4286.ln(RM_2(G))$   
 $C = -208.1414 + 100.0905.ln(RM_2(G))$   
 $LOGP = -1.2177 + 0.0817.ln(RM_2(G))$   
 $MP = 338.2667 - 14.9125.ln(RM_2(G))$   
 $CCS = 10.3761 + 23.0143.ln(RM_2(G))$

### 5.3 Reducible reciprocal randic index $RR(G)$

$MM = 120.8357 + 0.6734(RR(G))$	$MM = -0.0004(RR(G))^2 + 0.973(RR(G)) + 96.7815$
$XLOGP3 = 0.3582 - 0.005922(RR(G))$	$XLOGP3 = 0.0000(RR(G))^2 + 0.0001(RR(G)) - 0.123$
$C = 255.4439 + 0.8391(RR(G))$	$C = -0.0004(RR(G))^2 + 1.1469(RR(G)) + 230.722$
$LOGP = 0.4514 - 0.003701(RR(G))$	$LOGP = 0.0000(RR(G))^2 + 0.0087(RR(G)) - 0.5473$
$MP = 262.8292 - 0.1017(RR(G))$	$MP = 0.0000(RR(G))^2 - 0.1053(RR(G)) + 263.1149$
$CCS = 127.921 + 0.1565(RR(G))$	$CCS = -0.0002(RR(G))^2 + 0.2695(RR(G)) + 116.6876$

$MM = -399.3469 + 141.2089.ln(RR(G))$   
 $XLOGP3 = 3.9097 - 1.0407.ln(RR(G))$   
 $C = -283.9748 + 154.5721.ln(RR(G))$   
 $LOGP = 2.0468 - 0.5277.ln(RR(G))$   
 $MP = 336.1482 - 20.3144.ln(RR(g))$   
 $CCS = -6.5383 + 35.0199.ln(RR(G))$

## 5.4 Reducible reciprocal sum connectivity index $RSC(G)$

$$\begin{aligned}
 MM &= 81.6943 + 2.0309(RSC(G)) & MM &= -0.0006(RSC(G))^2 + 2.1967(RSC(G)) + 75.7318 \\
 XLOGP3 &= 0.6249 - 0.0172(RSC(G)) & XLOGP3 &= -0.0001(RSC(G))^2 + 0.0115(RSC(G)) - 0.4058 \\
 C &= 207.3057 + 2.525(RSC(G)) & C &= 0.0026(RSC(G))^2 + 1.8426(RSC(G)) + 231.8508 \\
 LOGP &= 0.5319 - 0.01001(RSC(G)) & LOGP &= -0.0002(RSC(G))^2 + 0.0311(RSC(G)) - 0.9455 \\
 MP &= 268.2718 + 0.4817(RSC(G)) & MP &= -0.0004(RSC(G))^2 - 0.1936(RSC(G)) + 264.3534 \\
 CCS &= 116.8003 + 0.4817(RSC(G)) & CCS &= -0.0007(RSC(G))^2 + 0.6722(RSC(G)) + 109.0213
 \end{aligned}$$

$$\begin{aligned}
 MM &= -479.3429 + 182.2592.\ln(RSC(G)) \\
 XLOGP3 &= 4.4663 - 1.3357.\ln(RSC(G)) \\
 C &= -389.4377 + 203.5928.\ln(RSC(G)) \\
 LOGP &= 2.2672 - 0.6632.\ln(RSC(G)) \\
 MP &= 350.0164 - 26.7631.\ln(RSC(G)) \\
 CCS &= -24.2619 + 44.7407.\ln(RSC(G))
 \end{aligned}$$

## 5.5 Reducible hyper first Zag-index $RHM_1(G)$

$$\begin{aligned}
 MM &= 162.7938 + 0.00754(RHM_1(G)) & MM &= 0.0000(RHM_1(G))^2 + 0.0149(RHM_1(G)) + 122.6824 \\
 XLOGP3 &= 0.3402 - 0.00008324(RHM_1(G)) & XLOGP3 &= 0.0000(RHM_1(G))^2 + 0.0001(RHM_1(G)) - 0.4118 \\
 C &= 308.3285 + 0.009365(RHM_1(G)) & C &= 0.0000(RHM_1(G))^2 + 0.0168(RHM_1(G)) + 267.8934 \\
 LOGP &= 0.5904 - 0.00005927(RHM_1(G)) & LOGP &= 0.0000(RHM_1(G))^2 + 0.0002(RHM_1(G)) - 0.5996 \\
 MP &= 254.006 - 0.0009953(RHM_1(G)) & MP &= 0.0000(RHM_1(G))^2 - 0.0021(RHM_1(G)) + 259.4173 \\
 CCS &= 141.1509 + 0.001613(RHM_1(G)) & CCS &= 0.0000(RHM_1(G))^2 + 0.0038(RHM_1(G)) + 124.5373
 \end{aligned}$$

$$\begin{aligned}
 MM &= -544.4697 + 96.6785.\ln(RHM_1(G)) \\
 XLOGP3 &= 5.6758 - 0.7905.\ln(RHM_1(G)) \\
 C &= -348.72 + 95.2908.\ln(RHM_1(G)) \\
 LOGP &= 3.3129 - 0.4423.\ln(RHM_1(G)) \\
 MP &= 315.7988 - 9.2112.\ln(RHM_1(G)) \\
 CCS &= -51.3892 + 24.7484.\ln(RHM_1(G))
 \end{aligned}$$

## 5.6 Reducible hyper first Zag-index $RHM_2(G)$

$$\begin{aligned}
 MM &= 210.4404 + 0.00007769(RHM_2(G)) & MM &= 0.0000(RHM_2(G))^2 + 0.0002(RHM_2(G)) + 178.4115 \\
 XLOGP3 &= -0.02199 - 9.748e-7(RHM_2(G)) & XLOGP3 &= 0.0000(RHM_2(G))^2 + 0(RHM_2(G)) - 0.3579 \\
 C &= 366.7098 + 0.00009706(RHM_2(G)) & C &= 0.0000(RHM_2(G))^2 + 0.0002(RHM_2(G)) + 326.4705 \\
 LOGP &= 0.4716 - 7.936e-7(RHM_2(G)) & LOGP &= 0.0000(RHM_2(G))^2 + 0(RHM_2(G)) - 0.2598 \\
 MP &= 248.8652 - 0.00001071(RHM_2(G)) & MP &= 0.0000(RHM_2(G))^2 + 0(RHM_2(G)) + 257.5147 \\
 CCS &= 155.0311 + 0.00001535(RHM_2(G)) & CCS &= 0.0000(RHM_2(G))^2 + 0(RHM_2(G)) + 145.1945
 \end{aligned}$$

$$\begin{aligned}
 MM &= -318.8317 + 54.8104.\ln(RHM_2(G)) \\
 XLOGP3 &= 3.4866 - 0.4186.\ln(RHM_2(G)) \\
 C &= -202.41 + 60.5619.\ln(RHM_2(G)) \\
 LOGP &= 1.9401 - 0.2215.\ln(RHM_2(G)) \\
 MP &= 336.6867 - 8.9537.\ln(RHM_2(G)) \\
 CCS &= 21.5129 + 13.0429.\ln(RHM_2(G))
 \end{aligned}$$

TABLE 15 Comparison table of correlation coefficient of logarithmic regression.

Indices	MM	XLOGP3	C	LOGP	MP	CCS
RM <sub>1</sub> (G)	0.9742	-0.3301	0.6346	-0.18	-0.3417	0.9861
RM <sub>2</sub> (G)	0.886	-0.177	0.5293	0.0358	-0.3129	0.8559
RR(G)	0.9697	-0.3168	0.6025	-0.1704	-0.3202	0.985
RSC(G)	0.9679	-0.3144	0.6137	-0.1656	-0.3264	0.9817
RHM <sub>1</sub> (G)	0.9306	-0.3373	0.5207	-0.2002	-0.2037	0.9173
RHM <sub>2</sub> (G)	0.9632	-0.326	0.6041	-0.183	-0.3603	0.981
RS(G)	0.9729	-0.4117	0.6375	-0.2583	-0.3225	0.9922
RF(G)	0.9742	-0.3466	0.6063	-0.1978	-0.3252	0.9901
RG <sub>1</sub> (G)	0.974	-0.3279	0.5996	-0.1831	-0.3157	0.9896
RG <sub>2</sub> (G)	0.9727	-0.336	0.5877	-0.1911	-0.3124	0.9906

### 5.7 Reducible sigma index RS(G)

$$\begin{aligned}
 MM &= 192.419 + 0.04944(RS(G)) & MM &= 0.0000(RS(G))^2 + 0.129(RS(G)) + 138.522 \\
 XLOGP3 &= 0.7616 - 0.0008381(RS(G)) & XLOGP3 &= 0.0000(RS(G))^2 + 0.0006(RS(G)) - 0.2409 \\
 C &= 335.0517 + 0.06534(RS(G)) & C &= 0.0000(RS(G))^2 + 0.1454(RS(G)) + 280.8399 \\
 LOGP &= 0.9285 - 0.0006116(RS(G)) & LOGP &= 0.0000(RS(G))^2 + 0.0007(RS(G)) + 0.0229 \\
 MP &= 246.1553 - 0.004883(RS(G)) & MP &= 0.0000(RS(G))^2 - 0.04(RS(G)) + 267.6473 \\
 CCS &= 146.93 + 0.01019(RS(G)) & CCS &= 0.0000(RS(G))^2 + 0.0276(RS(G)) + 133.0309
 \end{aligned}$$

$$\begin{aligned}
 MM &= -294.8106 + 93.1536.ln(RS(G)) \\
 XLOGP3 &= 4.4765 - 0.8895.ln(RS(G)) \\
 C &= -206.2213 + 107.5349.ln(RS(G)) \\
 LOGP &= 2.8294 - 0.5261.ln(RS(G)) \\
 MP &= 322.1681 - 13.6067.ln(RS(G)) \\
 CCS &= 27.9428 + 21.7919.ln(RS(G))
 \end{aligned}$$

### 5.8 Reducible forgotten index RF(G)

$$\begin{aligned}
 MM &= 169.4351 + 0.01312(RF(G)) & MM &= 0.0000(RF(G))^2 + 0.0267(RF(G)) + 132.5 \\
 XLOGP3 &= 0.3495 - 0.0001521(RF(G)) & XLOGP3 &= 0.0000(RF(G))^2 + 0.0002(RF(G)) - 0.487 \\
 C &= 312.626 + 0.01665(RF(G)) & C &= 0.0000(RF(G))^2 + 0.0329(RF(G)) + 268.6353 \\
 LOGP &= 0.5932 - 0.000108(RF(G)) & LOGP &= 0.0000(RF(G))^2 + 0.0003(RF(G)) - 0.5548 \\
 MP &= 253.8904 - 0.001809(RF(G)) & MP &= 0.0000(RF(G))^2 + 0.0068(RF(G)) + 266.2883 \\
 CCS &= 143.0493 + 0.002774(RF(G)) & CCS &= 0.0000(RF(G))^2 + 0.0067(RF(G)) + 128.9797
 \end{aligned}$$

$$\begin{aligned}
 MM &= -465.5766 + 95.8217.ln(RF(G)) \\
 XLOGP3 &= 4.9133 - 0.7692.ln(RF(G)) \\
 C &= -357.8937 + 105.0634.ln(RF(G)) \\
 LOGP &= 2.7509 - 0.4138.ln(RF(G)) \\
 MP &= 347.2109 - 13.9873.ln(RF(G)) \\
 CCS &= -19.4979 + 23.3651.ln(RF(G))
 \end{aligned}$$

## 5.9 Reducible first Gourava index $RG_1(G)$

$$\begin{array}{ll}
 MM = 160.963 + 0.03131(RG_1(G)) & MM = 0.0000(RG_1(G))^2 + 0.061(RG_1(G)) + 124.4735 \\
 XLOGP3 = 0.1775 - 0.0003094(RG_1(G)) & XLOGP3 = 0.0000(RG_1(G))^2 + 0(RG_1(G)) - 0.253 \\
 C = 304.648 + 0.03917(RG_1(G)) & C = 0.0000(RG_1(G))^2 + 0.0785(RG_1(G)) + 256.2399 \\
 LOGP = 0.4452 - 0.0002145(RG_1(G)) & LOGP = 0.0000(RG_1(G))^2 + 0.0005(RG_1(G)) - 0.4838 \\
 MP = 256.6838 - 0.004654(RG_1(G)) & MP = 0.0000(RG_1(G))^2 - 0.0109(RG_1(G)) + 263.8316 \\
 CCS = 140.455 + 0.006874(RG_1(G)) & CCS = 0.0000(RG_1(G))^2 + 0.0155(RG_1(G)) + 126.0261
 \end{array}$$

$$\begin{array}{l}
 MM = -463.4443 + 104.0763.ln(RG_1(G)) \\
 XLOGP3 = 4.5578 - 0.7904.ln(RG_1(G)) \\
 C = -346.22 + 112.8723.ln(RG_1(G)) \\
 LOGP = 2.4916 - 0.4162.ln(RG_1(G)) \\
 MP = 343.4525 - 14.7187.ln(RG_1(G)) \\
 CCS = -23.0011 + 25.8757.ln(RG_1(G))
 \end{array}$$

## 5.10 Reducible second Gourava index $RG_2(G)$

$$\begin{array}{ll}
 MM = 191.5845 + 0.0007926(RG_2(G)) & MM = 0.0000(RG_2(G))^2 + 0.0018(RG_2(G)) + 156.491 \\
 XLOGP3 = 0.09156 - 0.00000918(RG_2(G)) & XLOGP3 = 0.0000(RG_2(G))^2 + 0(RG_2(G)) - 0.3882 \\
 C = 342.2894 + 0.0009956(RG_2(G)) & C = 0.0000(RG_2(G))^2 + 0.0023(RG_2(G)) + 297.8582 \\
 LOGP = 0.4739 - 0.000006913(RG_2(G)) & LOGP = 0.0000(RG_2(G))^2 + 0(RG_2(G)) - 0.4004 \\
 MP = 251.5699 - 0.0001121(RG_2(G)) & MP = 0.0000(RG_2(G))^2 - 0.0005(RG_2(G)) + 262.0989 \\
 CCS = 149.7374 + 0.0001628(RG_2(G)) & CCS = 0.0000(RG_2(G))^2 + 0.0004(RG_2(G)) + 137.149
 \end{array}$$

$$\begin{array}{l}
 MM = -433.6237 + 73.2963.ln(RG_2(G)) \\
 XLOGP3 = 4.4809 - 0.5712.ln(RG_2(G)) \\
 C = -298.8449 + 78.0269.ln(RG_2(G)) \\
 LOGP = 2.5081 - 0.3063.ln(RG_2(G)) \\
 MP = 338.4152 - 10.2879.ln(RG_2(G)) \\
 CCS = -16.0368 + 18.2656.ln(RG_2(G))
 \end{array}$$

## 6 Statistical parameters in QSPR models

QSPR (Quantitative Structure-Property Relationship) modeling has emerged as a powerful tool in the field of computational chemistry to predict the properties and behaviors of chemicals and materials. In order to build reliable and accurate QSPR models, it is crucial to understand and utilize statistical parameters effectively. Statistical parameters play a crucial role in model selection. They help us identify the model that best captures the underlying relationship between chemical structure and property. By comparing statistical parameters across different models, we can determine which one is more accurate, precise, and reliable for making predictions. This ensures that we choose the most appropriate model for our specific needs. In QSPR modeling, several statistical parameters are used to assess model performance. These include correlation coefficient ( $R$  and  $R^2$ ), F-test value and  $p$ -value. By examining these parameters, we can determine whether our models are dependable and can consistently provide accurate predictions. This study delves into the discussion of eight statistical parameters for reducible topological indices within three regression models of QSPR. The variable "N" represents the size of the data set, which can vary from 9, 12, and 13. In correlation ( $r$ ), there are three possible cases. The first case occurs when  $r > 0$ , it means that as one term increases, the other term also tends to increase at the same time. It can be inferred that there is a powerful or undeviating connection between the two terms. The second case occurs when  $r < 0$ , implying that as one term increases, the other term tends to decrease. This indicates a negative or inverse connection between the two variables. The third case occurs when  $r = 0$ , meaning that no relationship is present between the two variables. The range of  $r$  values lies between  $-1$  and  $+1$  for the correlation coefficient. A value of  $+1$  indicates a perfect positive correlation, while a value of  $-1$  signifies a perfect negative correlation. In simpler terms, when the value of  $r$  is close to 1, such as 0.9 or 0.8, it suggests a strong and direct relationship between two variables. On the other hand, if the value of  $r$  is significantly far from 1, like 0.4, 0.3, or 0.2, it indicates a very weak association between the variables. The same principle applies when the value of  $r$  is nearly equal or far to  $-1$ . The correlation coefficient ( $R^2$ ) is a convenient tool for evaluating the QSPR models' quality.  $R^2 > 0.7$  is essential for a good QSPR model.

The F-test is a statistical test used to compare the variances of two populations. In simpler terms, it helps determine if the means of two groups are significantly different from each other, by comparing the variation within the groups to the variation between the



groups. The  $p$ -value associated with the F-test indicates the probability of observing the data if the null hypothesis (usually that there is no difference between the groups) is true. A low  $p$ -value (typically below 0.05) suggests that there is strong evidence against the null hypothesis, meaning the difference between the groups is unlikely to be due to random chance alone. The evaluation of our model's reliability is greatly enhanced by the utilization of two test values: the  $p$ -value and the F-test value. These values help us determine whether the relationship between properties and indices is significant or not.

In statistical analysis, a  $p$ -value  $\leq 0.05$  and an F-test value  $> 2.5$  are often considered indicators of a statistically significant or strong relationship. On the other hand, if the  $p$ -value is greater than 0.05 and the F-test value is less than 2.5, it suggests a lack of significance, indicating a weak relationship between properties and indices. By assessing the  $p$ -value and F-test value, we ensure the validity of our models and make informed decisions. The numerical values of 8 parameters, including  $p$ -values and F-test values, related to all the defined indices and properties for tuberculosis disease are contained in Tables 3–12.

## 7 Graphical analysis

Bar graphs are a popular and effective way to visually represent data. The use of rectangular bars of different lengths to show the magnitude of data points and make comparisons between different categories or groups for numerous drugs. By presenting data in a graphical format, bar graphs make it easier for people to understand and interpret information quickly. In bar graphs, the x-axis and the y-axis are commonly known as the two axes used. The horizontal axis (x-axis) represents the all reducible indices, while the vertical axis (y-axis) showcases the values of all correlation related to properties. These vertical or horizontal rectangles represent the magnitude or value of each category. The length or height of the bars is proportional to the data which are represent. In each graph, the squared symbol denotes data points associated with the linear, quadratic, and logarithmic model. Microsoft excel has been used to create these graphs. In Figure 7, there is a visual representation of the compared correlation graph showing the relationship between various reducible degree-based indices and characteristics of tuberculosis drugs. Every single graph was created using the data from Tables 13–15. Figure 7 reveals a clear visualization of the range of  $r$  values.

Figure 7A illustrates the comparison of correlation values with blue, orange, and grey bars representing the linear-model, quadratic-model, and logarithmic-models, respectively. The correlation values between reducible indices and molar mass range from 0.77 to 0.99. In Figure 7B, the inverse correlation values for the linear and quadratic models lie between  $-0.1$  and  $-0.6$ . However, only  $RG_2(G)$  does not exhibit an inverse relationship in the linear model, as its correlation value is 0.2256. On the other hand, the range of positive correlation values for the quadratic model lies between 0.1 and 0.7. The correlation  $r$  results for the three models in Figure 7C are all positive and range from 0.4 to 0.7. In Figure 7D, the inverse correlation values for the linear and quadratic models lie between  $-0.1$  and  $-0.5$ . However, only  $RG_2(G)$  does not show an inverse relationship in the linear and quadratic models, as its correlation values are 0.1487 and 0.0358. On the other hand, the range of positive correlation values ( $r$ ) for the quadratic model lies between 0.2 and 0.6. In Figure 7E, both the linear and quadratic models exhibit an inverse relationship due to their negative correlation values, which fall to  $-0.3$ . However, the quadratic model demonstrates positive correlation values ranging from 0.30 to 0.39. Lastly, in Figure 7F, all models exhibit a positive relationship, ranging from 0.6 to 0.9.

## 8 Results with discussion

The main objective of this portion is to discuss a QSPR analysis that focuses on ten specified reducible topological indices (TIs) and six distinct properties associated with tuberculosis disease. The mentioned properties comprise molar mass, XLOGP3, Complexity, LOGP, melting point, and collision cross section. The data for these properties has been sourced from the PubChem website and is presented in Table 2. Vertex partitions and edge partitions, which are degree-based TIs techniques, have been employed to compute the numerical values corresponding to the reducible molecular descriptors. The values obtained through calculation for all indices are present in Table 1. The association between reducible topological descriptors and physicochemical characteristics for anti-tuberculosis drugs has been determined in the three QSPR models that were examined. The statistical parameters were computed to obtain these results. Tables 3–12 presents the summarized values of eight statistical parameters that were calculated. All the results in the table include information on the  $p$ -values and F-test values for the ten reducible TIs associated with anti-tuberculosis drugs. As a result, some outcomes hold significant and strong value. For instance, Table 3 represents the statistical analysis for the reducible first Zagreb index. In this table, the linear model shows the F-test value with the highest and lowest significance values as (647.5515, 0.8791), with corresponding  $p$ -values of (0.0000, 0.3686). The correlation  $r$  values for this model are (0.9916,  $-0.272$ ). In the quadratic model, the F-test value is (139.5896, 0.6068) with a  $p$ -value of (0.0000, 0.566). The correlation  $r$  values for this model are (0.9981, 0.3446). Lastly, in the logarithmic model, the F-test value is (94.6245, 1.5637) with  $p$ -values of (0.0003, 0.5241). The correlation  $r$  values for this model are (0.9861,  $-0.3301$ ). This analysis provides valuable insights into the relationship between the reducible first Zagreb index and the different models. Figure 7, represents the bar graphs which are visually compared the  $r$  values between characteristics and reducible indices for anti-tuberculosis drugs. The graphs clearly illustrate the positive correlation  $r$  values observed in Figures 7A, C, F. Interestingly, Figures 7A, F exhibit a stronger positive correlation compared to Figure 7C. On the other hand, Figures 7B, D, E demonstrate that among the three models (linear, logarithmic, and quadratic), only the quadratic model displays a positive correlation value. Below the results with discussion for all indices that is based on Tables 13–15 are presented.

## 8.1 Linear regression model

By using Tables 3–13, we conclude that all of the defined reducible indices show a robust positive connection with molar mass within the range of  $r = 0.7$  to  $0.9$  with  $r^2 = 0.6$  to  $0.9$  having F-value, greater and lower value is (685.6292 and 20.1735) with  $p$ -value is less than  $0.05$ , as well as with collision cross section within the range of  $r = 0.6$  to  $0.9$  with  $r^2 = 0.4$  to  $0.9$  having F-value, greater and lower value is (452.0602 and 6.4616) with  $p$ -value is less than  $0.05$ . The complexity property demonstrates a significant correlation with  $RM_1(G)$  at  $r = 0.7074$ , with  $r^2 = 0.5004$  having F-value is 11.0191,  $RR(G)$  at  $r = 0.7016$  with  $r^2 = 0.4923$  and F-value is 10.6656, and  $RSC(G)$  at  $r = 0.7044$  and  $r^2 = 0.4962$  having F-value is 10.833. These three properties display a strong positive correlation with the indices. On the other hand, the remaining three properties, namely,  $XLOGP3$ ,  $LOGP$ , and  $MP$ , do not exhibit a strong correlation with any of the indices, as their correlation  $r$  values range from  $-0.1$  to  $-0.5$  with  $r^2 = -0.4$  to  $0.2$  having very low F-value and  $p$ -value which is greater than  $0.05$ . These values indicate an inverse relationship, and have a very low significance level. Only  $XLOGP3$  demonstrates a moderate inverse relationship with  $RS(G)$  at  $r = -0.6737$  with  $r^2 = 0.4538$  and F-value is 9.1398. Consequently, the properties of molar mass and collision cross section offer valuable insights into the relationship with all of the indices.

## 8.2 Quadratic regression model

Based on the results obtained from Tables 3–12, 14, it is clear that there exists a robust positive correlation between molar mass and all the specified indices, falling within the range of  $r = 0.8$  to  $0.9$  with  $r^2 = 0.7$  to  $0.9$  having F-value, highest and lowest value is (1448.3066 and 12.0901) with  $p$ -value is less than  $0.05$ . Additionally, a similar association is observed with collision cross section, ranging from  $r = 0.7$  to  $0.9$  and  $r^2 = 0.5, 0.8$  and  $0.9$  with F-value which is greater and lower is (567.848 and 3.97) having  $p$ -value which is less than  $0.05$ . These intriguing findings have a profound impact on the advancement of anti-tuberculosis drugs, as they provide remarkably valuable insights into the development process. Only one index, namely, the reducible sigma index at  $r = 0.7455$  with  $r^2 = 0.5558$  and F-value is 6.2551, demonstrates a good association with  $XLOGP3$ . The relationship between complexity with  $RM_1(G)$ ,  $RR(G)$ ,  $RSC(G)$ ,  $RS(G)$ ,  $RF(G)$ , and  $RG_1(G)$  demonstrates a strong correlation, with  $r$  values exceeding  $0.7$  with  $r^2 = 0.4$  to  $0.5$ . At a  $r = 0.6$ , this index demonstrates only a mediocre correlation with  $RM_2(G)$ ,  $RHM_1(G)$ ,  $RHM_2(G)$ , and  $RG_2(G)$ .  $LOGP$  show a mediocre correlation with only reducible sigma index at  $r = 0.6045$  with  $r^2 = 0.3655$ .

## 8.3 Logarithmic regression model

After analyzing the data presented in Tables 3–12, 15, we have reached the conclusion that all the reducible indices demonstrate a very high positive correlation with molar mass within the range of  $r = 0.8$  to  $0.9$  with  $r^2$  varies from  $0.7$  to  $0.9$  having F-value, greater and lower is (368.6357 and 57.5364) with  $p$ -value which is less than  $0.05$ , as well as with collision cross section within the same  $r$  and  $r^2$  range having F-value greater and lower is (479.3848, 39.7256) and  $p$ -values which is less than  $0.05$ . This analysis reveals a noteworthy relationship between the defined indices and both molar mass and collision cross section. The data clearly indicates a consistent and significant positive correlation within the specified range. The range of  $r = 0.6$  with  $r^2$  varies from  $0.4$  to  $0.3$  reveals a mediocre connection between the complexity property with six reducible indices, namely,  $RM_1(G)$ ,  $RR(G)$ ,  $RSC(G)$ ,  $RHM_2(G)$ ,  $RS(G)$ , and  $RF(G)$ . It can be observed that there is no strong correlation and significance level between  $LOGP$  and melting point across all defined indices due to the fact that the value of  $r$  falls between  $-0.1$  and  $-0.3$  with  $r^2 = 0.0$  to  $0.1$ .

## 9 Conclusion

Degree-based topological indices prove exceptionally valuable in modeling and characterizing the molecular structure of medications, allowing for the best prediction of physicochemical properties without conducting time-consuming experiments. These indices, derived from the principles of graph theory, serve as numerical descriptors for molecular structures. When analyzing drugs using QSPR, degree-based reducible topological indices serve as important tools for predicting drug properties by providing molecular descriptors. This knowledge is vital for the development of effective treatments and the optimization of drug design. The selected drugs such as Isoniazid, Pyrazinamide, Ethionamide, Levofloxacin, Amikacin, Ofloxacin, Cycloserine, 4-aminosalicylic acid, Ethambutol, Ciprofloxacin, Bedaquiline, Kanamycin and Streptomycin have been widely used in the treatment of tuberculosis. By calculating the degree-based reducible topological indices for these compounds, we can better understand their structural features and how they contribute to their pharmacological properties. Furthermore, we have developed a three-regression model to investigate the correlation between these calculated values and experimental properties, including molar mass,  $XLOGP3$ , Complexity,  $LOGP$ , melting point, and collision cross-section. Based on the statistical parameters and topological indices employed in the three QSPR model, we can conclude that both molar mass and collision cross section exhibit a highly significant association with all the specified indices. As an example, when examining the relationship between molar mass and collision cross section with  $RM_1(G)$  index, we observe varying correlation coefficients ( $r$  values) across different models. In

the linear model, the correlation coefficient ranges from 0.9916 to 0.9733. Meanwhile, the quadratic model yields a higher correlation coefficient, ranging from 0.9981 to 0.9947. Lastly, the logarithmic model shows a correlation coefficient ranging from 0.9742 to 0.9861. There is a good correlation between complexity and  $RM_1(G)$ ,  $RR(G)$ , and  $RSC(G)$  indices in the linear model at the value of  $r$  exceeds from 0.7. Similarly, in the quadratic model, the relationship between complexity and the  $RM_1(G)$ ,  $RR(G)$ ,  $RSC(G)$ ,  $RS(G)$ ,  $RF(G)$ , and  $RG_1(G)$  indices also demonstrates a good correlation, with  $r$  values exceeding from 0.7. Furthermore, within the three discussed models, it can be observed that there is no strong correlation or significance level between LOGP and melting point across all defined indices. This research indicates that theoretical assessment could allow chemists and other professionals in the pharmaceutical sector to predict the characteristics of tuberculosis medications without the need for conducting experiments. The analysis revealed a wide range of reducible topological indices, suggesting that alternative formulations of these drugs could potentially be used to treat a variety of disorders. The research also determined the correlation coefficient for several reducible topological indices, providing chemists with valuable information to develop new medications by combining existing ones with high correlations. We aspire for this research to create a pathway for future breakthroughs in drug design and therapy, ultimately resulting in enhanced results for individuals afflicted with this incapacitating illness. Before ending this paper, we present the following open problem for further examination.

## 9.1 Open problem

The analysis of QSPR focused on the correlation between different topological indices that is based on distance-based methodology, and physicochemical properties of anti-tuberculosis drugs through the use of exponential regression model and multi linear regression model.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## Author contributions

SA: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing–original draft, Writing–review and editing. NN: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing–original draft, Writing–review and editing. MR: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing–original draft, Writing–review and editing. LA: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing–original draft, Writing–review and editing.

## Funding

The authors declare that no financial support was received for the research, authorship, and/or publication of this article.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Natarajan A, Beena PM, Devnikar AV, Mali S. A systemic review on tuberculosis. *Indian J Tuberculosis* (2020) 67(3):295–311. doi:10.1016/j.ijtb.2020.02.005
- Jasmer RM, Nahid P, Hopewell PC. Latent tuberculosis infection. *New Engl J Med* (2002) 347(23):1860–6. doi:10.1056/nejmcp021045
- Ramirez-Lapausa M, Menendez-Saldana A, Noguero-Asensio A. Extrapulmonary tuberculosis: an overview. *Rev Esp Sanid Penit* (2015) 17(1):3–11. doi:10.4321/s1575-06202015000100002
- Sharma SK, Mohan A, Kohli M. Extrapulmonary tuberculosis. *Expert Rev Respir Med* (2021) 15(7):931–48. doi:10.1080/17476348.2021.1927718
- Mirzayev F, Viney K, Linh NN, Gonzalez-Angulo L, Gegia M, Jaramillo E, et al. World Health Organization recommendations on the treatment of drug-resistant tuberculosis, 2020 update. *Eur Respir J* (2021) 57(6):2003300. doi:10.1183/13993003.03300-2020
- Griffith DE, Kerr CM. Tuberculosis: disease of the past, disease of the present. *J Perianesthesia Nurs* (1996) 11(4):240–5. doi:10.1016/s1089-9472(96)80023-2
- Agyeman AA, Ofori-Asenso R. Tuberculosis-an overview. *J Public Health Emerg* (2017) 1(7):7–11. doi:10.21037/jphe.2016.12.08
- Hayat S, Imran M. On degree based topological indices of certain nanotubes. *J Comput Theor Nanoscience* (2015) 12(8):1599–605. doi:10.1166/jctn.2015.3935
- Siddiqui MK, Naeem M, Rahman NA, Imran M. Computing topological indices of certain networks. *J Optoelectronics Adv Mater* (2016) 18:884–92.
- Hayat S, Imran M. Computation of topological indices of certain networks. *Appl Math Comput* (2014) 240:213–28. doi:10.1016/j.amc.2014.04.091
- Bollobás B, Erdős P. Graphs of extremal weights. *Ars combinatoria* (1998) 50:225.
- Amic D, Bešlo D, Lucic B, Nikolic S, Trinajstić N. The vertex-connectivity index revisited. *J Chem Inf Comput Sci* (1998) 38(5):819–22. doi:10.1021/ci980039b
- Hu Y, Li X, Shi Y, Xu T, Gutman I. On molecular graphs with smallest and greatest zeroth-order general Randić index. *MATCH Commun Math Comput Chem* (2005) 54(2):425–34.
- Caporossi G, Gutman I, Hansen P, Pavlović L. Graphs with maximum connectivity index. *Comput Biol Chem* (2003) 27(1):85–90. doi:10.1016/s0097-8485(02)00016-5
- Li X, Gutman I. *Mathematical chemistry monographs no. 1*. Kragujevac, Macedonia: University of Kragujevac (2006).
- Kier LB, Hall LH. Molecular connectivity VII: specific treatment of heteroatoms. *J Pharm Sci* (1976) 65(12):1806–9. doi:10.1002/jps.2600651228
- Kier LB, Hall LH. *Molecular connectivity in structure-activity analysis* (1986). (No Title).
- Shafiei F. Relationship between topological indices and thermodynamic properties and of the monocarboxylic acids applications in QSPR. *Iranian J Math Chem* (2015) 6(1):15–28.
- Sepehri B. A review on created QSPR models for predicting ionic liquids properties and their reliability from chemometric point of view. *J Mol Liquids* (2020) 297:112013. doi:10.1016/j.molliq.2019.112013
- Hosamani SM, Kulkarni BB, Boli RG, Gadag VM. QSPR analysis of certain graph theoretical matrices and their corresponding energy. *Appl Math Nonlinear Sci* (2017) 2(1):131–50. doi:10.21042/amns.2017.1.00011
- Mondal S, Dey A, De N, Pal A. QSPR analysis of some novel neighbourhood degree-based topological descriptors. *Complex Intell Syst* (2021) 7:977–96. doi:10.1007/s40747-020-00262-0
- Miyao T, Kaneko H, Funatsu K. Inverse QSPR/QSAR analysis for chemical structure generation (from y to x). *J Chem Inf Model* (2016) 56(2):286–99. doi:10.1021/acs.jcim.5b00628
- Huang R, Mahboob A, Rasheed MW, Alam SM, Siddiqui MK. On molecular modeling and QSPR analysis of lyme disease medicines via topological indices. *The Eur Phys J Plus* (2023) 138(3):243. doi:10.1140/epjp/s13360-023-03867-9
- Hasani M, Ghods M. Topological indices and QSPR analysis of some chemical structures applied for the treatment of heart patients. *Int J Quan Chem* (2024) 124:e27234. doi:10.1002/qua.27234
- Mahboob A, Rasheed MW, Amin L, Hanif I. A study of novel molecular descriptors and quantitative structure–property relationship analysis of blood cancer drugs. *Eur Phys J Plus* (2023) 138(9):856. doi:10.1140/epjp/s13360-023-04499-9
- Mahboob A, Rasheed MW, Hanif I, Amin L, Alameri A. Role of molecular descriptors in quantitative structure-property relationship analysis of kidney cancer therapeutics. *Int J Quan Chem* (2024) 124:e27241. doi:10.1002/qua.27241
- Husin MN, Ariffin A. On the edge version of topological indices for certain networks. *Ital J Pure Appl Math* (2022) 47:550–64.
- Liu JB, Ali H, Shafiq MK, Dustigeer G, Ali P. On topological properties of planar octahedron networks. *Polycyclic Aromatic Comp* (2023) 43(1):755–71. doi:10.1080/10406638.2021.2022726
- Sardar MS, Ali MA, Ashraf F, Cancan M. On topological indices of double and strong double graph of silicon carbide Si<sub>2</sub> C<sub>3</sub>-I [p, q]. *Eurasian Chem Commun* (2023) 5:37–49. doi:10.22034/ecc.2023.356160.1519
- Raza Z, Akhter S, Shang Y. Expected value of first Zagreb connection index in random cyclooctatetraene chain, random polyphenyls chain, and random chain network. *Front Chem* (2023) 10:1067874. doi:10.3389/fchem.2022.1067874
- Nadeem I, Siddique S, Shang Y. Some inequalities between general randić-type graph invariants. *J Math* (2024) 2024:1–7. doi:10.1155/2024/8204742
- Gutman I, Trinajstić N. Graph theory and molecular orbitals. Total  $\phi$ -electron energy of alternant hydrocarbons. *Chem Phys Lett* (1972) 17(4):535–8. doi:10.1016/0009-2614(72)85099-1
- Islam SR, Pal M. Second Zagreb index for fuzzy graphs and its application in mathematical chemistry. *Iranian J Fuzzy Syst* (2023) 20(1):119–36. doi:10.22111/IJFS.2023.7350
- Mondal S, De N, Pal A. On neighborhood Zagreb index of product graphs. *J Mol Struct* (2021) 1223:129210. doi:10.1016/j.molstruc.2020.129210
- Randić M. Characterization of molecular branching. *J Am Chem Soc* (1975) 97(23):6609–15. doi:10.1021/ja00856a001
- Yuan M. On the Randić index and its variants of network data. *TEST* (2023) 1–25. doi:10.1007/s11749-023-00887-6
- Zhou B, Trinajstić N. On general sum-connectivity index. *J Math Chem* (2010) 47:210–8. doi:10.1007/s10910-009-9542-4
- Hussain M, ur Rehman A, Iqbal MS, Muhammad A. Computation of randić and sum-connectivity indices for H<sub>3</sub>BO<sub>3</sub> layer. *Structure's Derivatives* (2022). doi:10.21203/rs.3.rs-2124344/v1
- Carballosa W, Pestana D, Sigarreta JM, Tour E. Relations between the general sum connectivity index and the line graph. *J Math Chem* (2020) 58:2273–90. doi:10.1007/s10910-020-01180-9
- Shirdel GH, Rezapour H, Sayadi AM. *The hyper-Zagreb index of graph operations* (2013).
- Sarkar I, Nanjappa M. On second hyper-zagreb index of corona products related to R-graphs. *Int J Appl Math* (2023) 36(3):301. doi:10.12732/ijam.v36i3.1
- Hakeem A, Muhammad Katbar N, Muhammad F, Ahmed N. QSPR analysis of some important drugs used in heart attack treatment via degree-based topological indices and regression models. *Polycyclic Aromatic Comp* (2023) 1–10. doi:10.1080/10406638.2023.2262697
- Gutman I, Togan M, Yurttas A, Cevik AS, Cangul IN. Inverse problem for sigma index. *MATCH Commun Math Comput Chem* (2018) 79(2):491–508.
- Ascioglu M, Cangul IN. Sigma index and forgotten index of the subdivision and r-subdivision graphs. *Proc Jangjeon Math Soc* (2018) 21(3):375–83. doi:10.17777/pjms2018.21.3.375
- Havare OC. QSPR analysis with curvilinear regression modeling and topological indices. *Iranian J Math Chem* (2019) 10(4):331–41. doi:10.22052/ijmc.2019.191865.1448
- Furtula B, Gutman I. A forgotten topological index. *J Math Chem* (2015) 53(4):1184–90. doi:10.1007/s10910-015-0480-z
- Hosamani SM. A upper bound for the F<sub>1</sub>-index and its applications to fullerenes. *MATCH Commun Math Comput Chem* (2015) 74(1):97–101.
- De N, Nayeem SMA, Pal A. F-index of some graph operations. *Discrete Math algorithms Appl* (2016) 8(02):1650025. doi:10.1142/s1793830916500257
- Kulli VR. The Gourava indices and coincides of graphs. *Ann Pure Appl Math* (2017) 14(1):33–8. doi:10.22457/apam.v14n1a4
- Wang Y, Aslam A, Idrees N, Kanwal S, Iram N, Razaque A. On trees with given independence numbers with maximum Gourava indices. *Symmetry* (2023) 15(2):308. doi:10.3390/sym15020308